PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C07H 21/00, C07K 15/00 C12N 9/12, 15/15, 15/54 C12N 15/63, C12P 21/02	A1	(11) International Publication Number: WO 94/00469 (43) International Publication Date: 6 January 1994 (06.01.94)

(21) International Application Number: PCT/US93/06093

(22) International Filing Date: 25 June 1993 (25.06.93)

(30) Priority data:
07/905,600 26 June 1992 (26.06.92) US

(71) Applicant: IMMUNEX CORPORATION [US/US]; 51 University Street, Seattle, WA 98101 (US).

(72) Inventor: ZIEGLER, Steven, F. ; 2230 West Halladay Street, Seattle, WA 98199 (US).

(74) Agent: SEESE, Kathryn, A.; Immunex Corporation, 51 University Street, Seattle, WA 98101 (US).

(81) Designated States: AU, CA, FI, JP, KR, NO, NZ, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: NOVEL TYROSINE KINASE

(57) Abstract

A novel receptor protein tyrosine kinase named *ork* (Orphan receptor tyrosine kinase) is identified and characterized. cDNA encoding the *ork* protein is inserted into an expression vector for production of the protein via recombinant DNA technology. The *ork* cDNA, when transfected into COS-7 cells, encodes a 140Kd protein with *in vitro* kinase activity. The *ork* gene is expressed predominantly in placenta and lung, with lower levels in umbilical vein endothelial cells, brain and kidney.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**MR**	Mauritania
AU	Australia	**GA**	Gabon	**MW**	Malawi
BB	Barbados	**GB**	United Kingdom	**NE**	Niger
BE	Belgium	**GN**	Guinea	**NL**	Netherlands
BF	Burkina Faso	**GR**	Greece	**NO**	Norway
BG	Bulgaria	**HU**	Hungary	**NZ**	New Zealand
BJ	Benin	**IE**	Ireland	**PL**	Poland
BR	Brazil	**IT**	Italy	**PT**	Portugal
BY	Belarus	**JP**	Japan	**RO**	Romania
CA	Canada	**KP**	Democratic People's Republic of Korea	**RU**	Russian Federation
CF	Central African Republic			**SD**	Sudan
CG	Congo	**KR**	Republic of Korea	**SE**	Sweden
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovak Republic
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LU**	Luxembourg	**TD**	Chad
CS	Czechoslovakia	**LV**	Latvia	**TG**	Togo
CZ	Czech Republic	**MC**	Monaco	**UA**	Ukraine
DE	Germany	**MG**	Madagascar	**US**	United States of America
DK	Denmark	**ML**	Mali	**UZ**	Uzbekistan
ES	Spain	**MN**	Mongolia	**VN**	Viet Nam
FI	Finland				

NOVEL TYROSINE KINASE

BACKGROUND

The ability of cells to respond to environmental cues is in large part due to the interaction of cell-surface receptors with external stimuli. A number of factors that interact with receptors have been identified, among which are receptor-binding proteins that are soluble, membrane-bound, or exist in both forms.

One class of receptors, the receptor tyrosine kinases (RTKs), has been intensively studied and shown to be crucial to the growth and differentiation of a variety of cell types (Yarden and Ullrich, *Ann. Rev. Biochem.* 57:433-478, 1988). Tyrosine kinases are enzymes that catalyze the phosphorylation of tyrosine residues. Tyrosine phosphorylation is associated with signal-transduction across the cellular plasma membrane. The protein-tyrosine kinase family can be grouped into two very broad families: the above-mentioned RTKs, which are, or are intimately associated with, membrane-spanning growth factor receptors; and those that are associated with the membrane but lack a transmembrane sequence (Yarden and Ullrich, 1988, *supra.*).

The RTKs can be further divided into five subgroups on the basis of structural similarities in their extracellular domains and the organization of the tyrosine kinase catalytic region in their cytoplasmic domains. Subgroups I (epidermal growth factor (EGF) receptor-like), II (insulin receptor-like) and the *eph/eck* family contain cysteine-rich sequences (Hirai *et al., Science* 238:1717-1720, 1987; Yanden and Ullrich, *Ann. Rev. Biochem*, 57,443-478, 1988; Lindberg and Hunter, *Mol. Cell. Biol.* 10:6316-6324, 1990). The functional domains of the kinase region of these three classes of RTKs are encoded as a contiguous sequence (Hanks *et al., Science* 241, 42-52, 1988). Subgroups III (platelet- derived growth factor (PDGF) receptor-like) and IV (the fibroblast growth factor (FGF) receptors) are characterized as having immunoglobulin (Ig)-like folds in their extracellular domains, as well as having their kinase domains divided in two parts by a variable stretch of unrelated amino acids (Yanden and Ullrich, 1988; Hanks *et al.*, 1988).

While all members of the RTK family share a related cytoplasmic catalytic domain, the extracellular, ligand-binding domains of these receptors have adapted patchwork structure utilizing several structural motifs. The variablility in the structure of the ligand-binding domains of the RTKs almost certainly reflects the diversity of the ligands for these receptors (Ullrich and Schlessinger, *Cell* 61, 243-254, 1990). These ligands range from relatively small, soluble peptides to cell surface proteins that

themselves resemble receptors. Examples of ligands bound by certain members of the RTK family are polypeptide growth factors and hormones.

To deal with this diversity of ligands the RTKs have evolved extracellular domains that are a composite of several structural motifs. For example, the extracellular domain of the *axl/ark* gene contains both Ig-domains and fibronectin type III (FNIII) repeats (O'bryan et al., *Mol. Cell. Biol.* 11:5016-5031, 1991; Rescigno et al., *Oncogene* 6:1909-1913, 1991), while members of the *eph* family have those two motifs separated by a (non-EGF-like) cysteine-rich domain (Hirai et al., *Science* 238:1717-1720, 1987; Lindberg et al., *Mol. Cell. Biol.* 10:6316-6324, 1990; Lhotak et al., *Mol. Cell. Biol.* 11:2496-2502, 1991; Chan and Watt, *Oncogene* 6:1057-1061,1991). This diversity strongly suggests that this family of receptors evolved by accumulating the structural motifs needed for ligand binding and combining these motifs with a conserved catalytic domain.

In view of the role tyrosine kinases play in cell growth and differentiation, as well as signal transduction, isolation of novel tyrosine kinases enables one to study such biological processes. Identifying tyrosine kinases and their ligands also permits exploration of methods for inhibiting or enhancing the interaction thereof, depending on the desired biological effect.

SUMMARY OF INVENTION

The present invention provides a novel receptor (a protein tyrosine kinase), isolated DNA encoding the tyrosine kinase, recombinant expression vectors containing the isolated DNA, and host cells transformed with the recombinant vector. Also provided is a method for producing the novel protein by cultivating the transformed host cells under conditions that promote expression of the tyrosine kinase, and recovering the expressed tyrosine kinase.

BRIEF DESCRIPTION OF FIGURES

Figure 1 presents a nucleotide sequence and deduced amino acid sequence of human *ork* cDNA. The nucleotide sequence is derived from two overlapping cDNA clones isolated from placental cDNA libraries. The initiating methionine codon and the transmembrane region are each underlined. The two cysteines of the immunoglobulin domain are boxed, as are the three EGF-like repeats. Brackets enclose the region amplified by PCR using primers based on conserved kinase domain sequences, as described in example 1. Numbering of nucleotides is in the left margin, while numbering of amino acids is on the right. The signal peptide comprises amino acids 1-18, with the threonine residue at position 19 being the first amino acid of the mature protein. The extracellular domain comprises amino acids 19-745, and the cytoplasmic domain comprises amino acids 773-1124.

Figure 2 presents an analysis of the EGF repeats in the *ork* sequence. The three EGF-repeats in the *ork* sequence were aligned with those in the *tan*-1 and laminin B1 genes. The conserved cysteine residues are boxed and other conserved residues (conserved in two *ork* repeats and at least one other repeat) are marked with an asterisk.

Figure 3 presents the results of studies described in example 2 below. The human *ork* cDNA encodes a 140 Kd phosphoprotein with *in vitro* tyrosine kinase activity. **A.** COS cells were transfected with the *ork* cDNA (*ork*), or a control plasmid without cDNA insert (mock), metabolically labelled with ^{35}S-met/cys, and immunoprecipitated with the P1 rabbit antiserum, raised against the carboxy-terminal 21 amino acids of the predicted *ork* amino acid sequence (P1), or with pre-immune serum (C). The position of P140^{ork} is indicated, as are the positions of size standards. **B.** COS cells were transfected and immunoprecipitated as described above. The immunoprecipitates were incubated with ^{32}P-ATP and separated by SDS-PAGE. The positions of P140^{ork} and size standards are indicated. **C.** The gel from Panel B was transferred to a membrane, exposed to X-ray film and the band corresponding to P140^{ork} was excised. The filter slice was boiled in 6N HCl and eluted material was separated by two-dimensional electrophoresis. The positions of standards for phosphoserine, phosphothreonine and phosphotyrosine, as well as the origin, are indicated. **D.** COS cells were transfected and immunoprecipitated as described above. The immunoprecipitates were separated by SDS-PAGE, transferred to a membrane and incubated with an anti-phosphotyrosine antibody. The positions of size standards are indicated.

Figure 4 shows the results of a study of expression of *ork* mRNA in human tissues, as described in example 3 below. Two micrograms of poly(A)-containing RNA from the indicated tissues was hybridized with an ^{32}P-UTP-labelled RNA probe from the *ork* cDNA. The positions of size standards are marked.

Figures 5a and 5b depict a comparison of the amino acid sequences of *ork* (top lines) and *tie* (bottom lines), the latter being a tyrosine kinase of the same subgroup as *ork*. The sequence comparison was generated by the GAP computer program described. Identical amino acids are indicated with a line; conservative or similar amino acid changes are indicated with two dots or one dot, respectively. The percent similarity and percent identity for the *ork* and *tie* sequences were found to be 64.5% and 47.5%, respectively.

Figure 6 presents a comparison of the amino acid sequence encoded by a murine *ork* clone with the corresponding portion of a human *ork* protein.

DETAILED DESCRIPTION OF INVENTION

The present invention provides a novel receptor protein tyrosine kinase, isolated DNA encoding the tyrosine kinase, recombinant expression vectors containing the isolated DNA, and host cells transformed with the recombinant vector. A method for producing the novel protein involves cultivating the transformed host cells under conditions that promote expression of the tyrosine kinase, and recovering the expressed tyrosine kinase from the cell culture.

The novel protein is designated herein as *ork* (*o*rphan *r*eceptor tyrosine *k*inase). An *ork* cDNA isolated from human placenta and transfected into COS-7 cells encodes a 140Kd protein with the following combination of structural motifs in its extracellular domain: an immunoglobulin (Ig)-like domain followed by three epidermal growth factor (EGF)-like cysteine-rich repeats, which in turn are followed by three fibronectin type III (FNIII) repeats. Of the human tissues tested (example 3 below), the *ork* gene is expressed predominately in placenta and lung, with lower levels in umbilical vein endothelial cells, brain and kidney.

Human *ork* is within the scope of the present invention, as are *ork* proteins derived from other mammalian species. As used herein, the term "*ork*" includes membrane-bound proteins (comprising a cytoplasmic domain, a transmembrane region, and an extracellular domain) as well as truncated proteins that retain the desired biological properties. Such truncated proteins include, for example, soluble *ork* comprising only the extracellular (ligand binding) domain.

Using a polymerase chain reaction-based approach we have isolated *ork* cDNA and characterized this novel receptor protein tyrosine kinase. As described in example 1 below, degenerate oligonucleotide probes based on certain sequences that are conserved within the kinase domain of RTKs were used as primers in a PCR reaction. Single-stranded cDNA derived from human placental poly(A)$^+$ RNA was employed as the template. The PCR reaction products were inserted into a cloning vector and the nucleotide sequence of the cDNA inserts in a number of the resulting recombinant vectors was determined. A clone containing a cDNA insert of about 200 basepairs with a novel DNA sequence was identified.

The cDNA insert was excised, radiolabeled, and used to probe an oligo-dT-primed human placental cDNA library. A hybridizing clone comprising a 2.3 Kb insert was identified, sequenced, and found to be a partial clone encoding a C-terminal *ork* fragment. The cDNA insert was excised, radiolabeled, and used to probe a random-primed human placental cDNA library. A clone comprising a cDNA insert about 4.0 Kb in length that contained an entire coding region for the novel protein, as well as 5' and 3' untranslated sequences, was isolated. The 4.0 kb *ork* cDNA in plasmid

pBLUESCRIPT®SK in *E. coli* strain DH5α was deposited with the American Type Culture collection on May 28, 1992 and was given accession number ATCC 69003. The strain deposit was made under the terms of the Budapest Treaty.

A DNA sequence for a human *ork* cDNA is presented in Figure 1, along with the amino acid sequence encoded thereby. Human *ork* DNA and amino acid sequences identical to those of Figure 1 are presented in SEQ ID NO:1 and SEQ ID NO:2. Several features of the *ork* amino acid sequence are noteworthy. The extracellular domain is a patchwork of three structural motifs. Between amino acid residues 211 and 340 are three copies of an EGF-like cysteine repeat (Davis, *New Biologist* 2:410-419, 1991). The EGF-like repeats in the *ork* gene (boxed in Figure 1) differ from the consensus motif in that they have 8 cysteines each instead of 6. In this regard they are closely related to those in the *tie* gene (Partanen et al., *Mol. Cell. Biol.*, 12:1698-1707, 1992). The proteins with the next most closely related cysteine repeats are laminin B1, laminin B2 (Sasaki et al., *J. Biol. Chem* 263:16536-16544, 1988) and *TAN*, the human homolog of the Drosophila *notch* gene (Ellison et al, *Cell* 66:649-661). The three EGF-repeats in the *ork* sequence were aligned with those in the *tan*-1 and laminin B1 genes in Figure 2. In each case the cysteines in the repeat were aligned.

Amino terminal to the EGF repeats are paired cysteines (boxed in Figure 1) indicative of Ig-like domains. This Ig domain is most similar to the C2 type as described by Williams and Barclay, *Ann. Rev. Immunol.* 6:381-405, 1988. A second pair of cysteine residues follow the EGF repeats; however the additional amino acid residues required to generate a proper Ig-fold are missing. At the membrane proximal region of the extracellular domain, between amino acids 440 and 733, are three repeats of the fibronectin type III motif. Thus, the extracellular domain of *ork* is a structural mosaic consisting of repeats of at least three different functional motifs.

Amino acid residues 746-772 are hydrophobic in nature and likely serve as a transmembrane domain. The cytoplasmic domain contains all the amino acid sequence hallmarks of a tyrosine kinase (Hanks et al., *Science* 241:42-52, 1988). A comparison of the *ork* amino acid sequence with known protein tyrosine kinases showed *ork* to be a member of the RTK family.

Within the receptor tyrosine kinase family, the *ork* and above-mentioned *tie* genes form their own subgroup. The *tie* gene (described by Partanen et al., *supra*) has an extracellular domain that, like *ork*, comprises (from N- to C-terminus) an Ig-like domain followed by three EGF-like domains, an incomplete Ig-like domain, and three fibronectin type III repeats next to the transmembrane region. Additionally, a cDNA form lacking the first of the three epidermal growth factor homology domains was isolated by Partanen et al., *supra*, suggesting that alternative splicing creates different

tie-type receptors. COS-7 cells transfected with a *tie* cDNA expression vector produced a glycosylated protein of 117 kDa.

A comparison of the amino acid sequences of *ork* and *tie* is presented in Figures 5a and 5b, which shows them to be distinct proteins. The comparison presented in Figures 5a and 5b was generated using the GAP computer program, which is described in detail below. The percent similarity and percent identity of the *ork* and *tie* amino acid sequences was found to be 64.5% and 47.5%, respectively. The RTK subgroup containing *ork* and *tie* is characterized by a unique arrangement of structural motifs in the proteins' extracellular domain, as described above. There are also several features in the cytoplasmic kinase domain worth noting. There is a short amino acid insert between the two parts of the kinase domains (Fig. 1). This insert is similar in size to that seen in the *ret* gene (Takahashi and Cooper, *Mol. Cell. Biol.* 7:1378-1385, 1987). However, unlike the insert sequences in other RTK subgroups, the insert in the *ork* and *tie* genes contains no tyrosine residues. Several tyrosine residues are found in the cytoplasmic tail. As phosphorylation of tyrosine in either the kinase insert or the cytoplasmic tail is required for association of the RTK molecule with putative signal transduction molecules, this feature of the *ork* and *tie* genes mostly closely resembles the EGFR subfamily where phosphorylated tyrosine residues in the cytoplasmic tail have been shown to associate with other proteins such as phospholipase C-γ (Margolis et al., *Cell* 57:1101-1107, 1989).

One embodiment of the present invention provides an *ork* protein comprising an extracellular domain that comprises (from N- to C-terminus) an immunoglobin-like domain, three EGF-like cysteine-rich repeats, and three FNIII repeats, wherein the amino acid sequence of said extracellular domain is at least 80% identical to the sequence shown as amino acids 19-745 of SEQ ID NO:1. The extracellular domain amino acid sequence preferably is at least 90% identical to the extracellular domain sequence of SEQ ID NO:1. The percent identity between two amino acid sequences may be determined by using the GAP computer program available from the University of Wisconsin and described in detail below. Such *ork* proteins include those having a transmembrane region and a cytoplasmic domain (or a portion thereof) in addition to the extracellular domain. Such proteins also include those in which a sequence containing a pair of cysteines is positioned between the last EGF-like repeat and the first FN III repeat. Such a sequence (a stretch of amino acids resembling an incomplete Ig-like domain) is found in the native human *ork* protein described above. Since this sequence is not believed to function as an Ig-like domain, it is possible that *ork* proteins lacking some or all of this partial Ig-like domain (amino acids 341-439) will retain the ability to bind a ligand. The *ork* proteins when initially synthesized may comprise a signal peptide as well.

The expression pattern of the *ork* gene suggests that it is predominately expressed in endothelial cells. The structure of the extracellular domain, with the three EGF repeats, is especially intriguing in light of this. Several cell adhesion molecules, including endothelial-leukocyte adhesion molecule-1 (ELAM-1) contain EGF repeats. ELAM-1 is expressed on the surface of activated endothelial cells and is involved in the attachment of neutrophils at sites of inflammation (Bevilacqua et al., *Science* 243:1150-1165, 1989; Siegelman et al., *Cell* 61:611-622, 1990). Several adhesion molecules also contain Ig and FNIII repeats as well. This suggests a possible role for the *ork* gene product in the communication between endothelial cells and leukocytes at sites of inflammation. cDNA encoding an *ork* polypeptide may be isolated from other mammalian species by procedures analogous to those employed in isolating the human *ork* clone. For example, a cDNA library derived from another mammalian species may be substituted for the human cDNA library that was screened using the degenerate probes in example 1 below. Alternatively (and preferably), the human *ork* cDNAs isolated in example 1 are labeled and used as probes to screen mammalian cDNA or genomic libraries using cross-species hybridization techniques. The probe may be derived from the coding region of the above-described 2.3kb or 4.0 kb human *ork* cDNAs.

Murine *ork* cDNA was identified by cross-species hybridization. A murine cDNA library was screened using the 2.3 Kb human *ork* cDNA as a probe, as described in example 5.

Cell types from which cDNA and genomic libraries may be prepared include those in which *ork* RNA expression was detected in example 3. mRNAs isolated from various cell lines can be screened by Northern hybridization to determine additional suitable sources of mammalian *ork* mRNA for use in cloning an *ork* gene. Nucleic acid from mammalian sources that include but are not limited to murine, bovine, porcine, and primate, may be screened to identify *ork* genes.

In addition to the membrane-bound full length protein depicted in SEQ ID NO:1, the present invention provides soluble forms of the *ork* protein. "Soluble *ork*" as used in the context of the present invention refers to polypeptides that contain all or part of the extracellular region of a native *ork* and that, due to the absence of a transmembrane region that would cause retention of the polypeptide on a cell membrane, are secreted upon expression. Fragments of the extracellular domain may be employed as long as the fragment possesses the desired biological activity (e.g., binding to an anti-*ork* antibody or to the ligand for *ork*). Soluble *ork* may also include part of the transmembrane region or part of the cytoplasmic domain or other sequences, provided that the soluble *ork* protein is capable of being secreted. Preferred soluble *ork* polypeptides include the signal sequence and entire extracellular domain (amino acids 1

to 745 of SEQ ID NO:1) or lack the signal sequence but contain the entire extracellular domain (amino acids 19 to 745 of SEQ ID NO:1).

Soluble *ork* polypeptides may be identified (and distinguished from their non-soluble membrane-bound counterparts) by separating intact cells which express the protein in question from the culture medium, e.g., by centrifugation, and assaying the medium (supernatant) for the presence of *ork*. The presence of *ork* in the medium indicates that the protein was secreted from the cells and thus is a soluble form. Soluble *ork* may be naturally-occurring forms of these proteins, such as those resulting from alternative splicing events. Alternatively, soluble fragments of *ork* proteins may be produced by recombinant DNA technology or otherwise isolated, as described below.

The use of soluble forms of *ork* is advantageous for certain applications. Purification of the proteins from recombinant host cells is facilitated, since the soluble proteins are secreted from the cells. The smaller soluble fragments may be advantageous for use in certain *in vitro* assays. The soluble *ork* polypeptides may be employed to competitively bind the ligand *in vivo*, thus inhibiting signal transduction activity via endogenous cell surface bound *ork* proteins. Further, soluble proteins are generally more suitable for intravenous administration and may exert their desired effect (e.g., binding a ligand) in the bloodstream.

Truncated *ork* proteins, including soluble polypeptides, may be prepared by any of a number of conventional techniques. In the case of recombinant proteins, a DNA fragment encoding a desired *ork* fragment may be subcloned into an expression vector. A desired DNA sequence may be chemially synthesized using known techniques. DNA fragments also may be produced by restriction endonuclease digestion of a full length cloned DNA sequence, and isolated by electrophoresis on agarose gels. Linkers containing restriction endonuclease cleavage site(s) may be employed to insert the desired DNA fragment into an expression vector, or the fragment may be digested at cleavage sites naturally present therein.

In another approach, enzymatic treatment (e.g., using Bal 31 exonuclease) may be employed to delete terminal nucleotides from a DNA fragment to obtain a fragment having a particular desired terminus. Among the commercially available linkers are those that can be ligated to the blunt ends produced by Bal 31 digestion, and which contain restriction endonuclease cleavage site(s). Alternatively, oligonucleotides that reconstruct the N- or C-terminus of a DNA fragment to a desired point may be synthesized. The oligonucleotide may contain a restriction endonuclease cleavage site upstream of the desired coding sequence and position an initiation codon (ATG) at the N-terminus of the coding sequence. The well known polymerase chain reaction procedure also may be employed to amplify a DNA sequence encoding a desired

protein fragment. 3' and 5' oligonucleotide primers that anneal to the *ork* DNA at the termini of a desired fragment are employed in the PCR reaction which is conducted using any suitable procedure, such as those described in Sarki et al., *Science* 239:487 (1988); in *Recombinant DNA Methodology*, Wu et al., eds., Academic Press Inc., San Diego (1989), pp. 189-196; and in *PCR Protocols: A Guide to Methods and Applications*, Innis et al., eds., Academic Press, Inc. (1990). An example of a suitable PCR procedure is as follows. All temperatures are in degrees centigrade. The following PCR reagents are added to a 0.5 ml Eppendorf microfuge tube: 10 µl of 10X PCR buffer (500 mM KCl, 100 mM Tris-HCl, pH 8.3 at 25°C, 25 mM $MgCl_2$, and 1 mg/ml gelatin) (Perkin-Elmer Cetus, Norwalk, CN), 8µl of a 2.5 mM solution containing each dNTP (2 mM dATP, 2 mM dCTP, 2 mM dGTP and 2 mM dTTP), 2.5 units (0.5 µl of standard 5000 units/ml solution) of *Taq* DNA polymerase (Perkins-Elmer Cetus), 1 ng of template DNA, 100 picomoles of each of the oligonucleotide primers, and water to a final volume of 100 µl. The final mixture is then overlaid with 100 µl parafin oil. PCR is carried out using a DNA thermal cycler (Ericomp, San Diego, CA). The template is denatured at 94° for 5 minutes and PCR is carried out for 25 cycles of amplification using a step program (denaturation at 94°, 1.5 minutes; annealing at 60°, 1 minute; extension at 72°, 1 minute).

The present invention also provides full length *ork* protein or antigenic fragments thereof that can act as immunogens to generate antibodies specific to the *ork* immunogens. Monoclonal antibodies specific for *ork* or antigenic fragments thereof are prepared by prodedures that include those described in Example 4. The above-described procedures for producing *ork* fragments may be employed in producing *ork* fragments for use as immunogens.

Expression of Recombinant *ork* Proteins

The present invention provides recombinant expression vectors to express DNA encoding the *ork* proteins of the present invention. The inventive recombinant expression vectors are replicable DNA constructs which contain a synthetic or cDNA-derived DNA sequence encoding an *ork* protein, operably linked to suitable transcriptional or translational regulatory elements. Examples of genetic elements having a regulatory role in gene expression include transcriptional promoters, operators or enhancers, a sequence encoding suitable mRNA ribosomal binding sites, and appropriate transcription and translation initiation and termination sequences. The ability to replicate in a host, usually conferred by an origin of replication, and a selection gene to facilitate recognition of transformants may additionally be incorporated. The regulatory elements employed in the expression vectors are generally

derived from mammalian, microbial, viral, or insect genes. Expression vectors derived from retroviruses also may be employed.

DNA regions are operably linked when they are functionally related to each other. A DNA sequence encoding *ork* is said to be operably linked to one or more of the above-described regulatory elements when the *ork* DNA sequence is transcribed, or the resulting mRNA is translated, under the control of the regulatory element(s).

Transformed host cells are cells which have been transformed or transfected with foreign DNA using recombinant DNA techniques. In the context of the present invention, the foreign DNA includes a sequence encoding the inventive *ork* protein. Host cells may be transformed for purposes of cloning or amplifying the foreign DNA, or may be transformed with an expression vector for production of the fusion protein under the control of appropriate promoters. Suitable host cells include prokaryotes, yeast, or higher eukaryotic cells. Appropriate cloning and expression vectors for use with bacterial, fungal, yeast, and mammalian cellular hosts are described by Pouwels et al. (*Cloning Vectors: A Laboratory Manual*, Elsevier, New York, 1985), the relevant disclosures of which is hereby incorporated by reference. Cell-free translation systems could also be employed to produce fusion protein using RNAs derived from the DNA constructs of the present invention.

Prokaryotes include gram negative or gram positive organisms. Prokaryotic expression vectors generally comprise one or more phenotypic selectable markers, for example a gene encoding proteins conferring antibiotic resistance or supplying an autotrophic requirement, and an origin of replication recognized by the host to ensure amplification within the host. Examples of suitable prokaryotic hosts for transformation include *E. coli*, bacilli such as *Bacillus subtilis*, *Salmonella typhimurium*, and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

Useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well-known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and pGEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. *E. coli* is typically transformed using derivatives of pBR322, a plasmid derived from an *E. coli* species (Bolivar et al., *Gene* 2:95, 1977). pBR322 contains genes for ampicillin and tetracycline resistance, providing simple means for identifying transformed cells.

Promoters commonly used in recombinant microbial expression vectors include the ß-lactamase (penicillinase) and lactose promoter system (Chang et al., *Nature*

275:615, 1978; and Goeddel et al., *Nature 281*:544, 1979), the tryptophan (trp) promoter system (Goeddel et al., *Nucl. Acids Res. 8*:4057, 1980; and EPA 36,776) and tac promoter (Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, p. 412, 1982). A particularly useful bacterial expression system employs the phage λ P_L promoter and cI857ts thermoinducible repressor. Plasmid vectors available from the American Type Culture Collection which incorporate derivatives of the λ P_L promoter include plasmid pHUB2, resident in *E. coli* strain JMB9 (ATCC 37092) and pPLc28, resident in E. coli RR1 (ATCC 53082).

The recombinant *ork* protein may also be expressed in yeast hosts, preferably from *Saccharomyces* species, such as *S. cerevisiae*. Yeast of other genera such as *Pichia* or *Kluyveromyces* may also be employed. Yeast vectors will generally contain an origin of replication from the 2μm yeast plasmid or an autonomously replicating sequence (ARS), a promoter, DNA encoding the *ork* protein, sequences for polyadenylation and transcription termination and a selection gene. Yeast vectors may include origins of replication and selectable markers permitting transformation of both yeast and *E. coli*, e.g., the ampicillin resistance gene of *E. coli* and the S. *cerevisiae* trp1 gene. The trp 1 gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, and a promoter derived from a highly expressed yeast gene to induce transcription of a structural sequence downstream. The presence of the trp1 lesion in the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan.

Suitable promoter sequences in yeast vectors include the promoters for metallothionein, 3-phosphoglycerate kinase (Hitzeman et al., *J. Biol. Chem. 255*:2073, 1980) or other glycolytic enzymes (Hess et al., *J. Adv. Enzyme Reg. 7*:149, 1968; and Holland et al., *Biochem. 17*:4900, 1978), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase and glucokinase. Examples of suitable vectors and promoters for use in yeast expression are further described in R. Hitzeman et al., EPA 73,657.

Preferred yeast vectors can be assembled using DNA sequences from pBR322 for selection and replication in *E. coli* (Amp^r gene and origin of replication) and yeast DNA sequences including a glucose-repressible ADH2 promoter and α–factor secretion leader. The ADH2 promoter has been described by Russell et al. (*J. Biol. Chem. 258*:2674. 1982) and Beier et al., (*Nature 300*:724, 1982). Advantageously, a DNA segment encoding a leader sequence functional in yeast is operably linked to the 5' end of the DNA encoding the *ork* protein. The encoded leader peptide promotes secretion of the *ork* protein from the host cell and is generally cleaved from the *ork*

protein upon secretion. As one example, the yeast α-factor leader, which directs secretion of heterologous proteins, can be inserted between the promoter and the structural gene to be expressed. See, e.g., Kurjan et al., *Cell 30*:922, 1982; and Bitter et al., *Proc. Natl. Acad. Sci. USA 81*:5330, 1984. The leader sequence may be modified to contain, near its 3' end, one or more useful restriction sites to facilitate fusion of the leader sequence to foreign genes.

Suitable yeast transformation protocols are known to those of skill in the art. An exemplary technique is described by Hinnen et al., *Proc. Natl. Acad. Sci. USA 75*:1929, (1978), selecting for Trp^+ transformants in a selective medium consisting of 0.67% yeast nitrogen base, 0.5% casamino acids, 2% glucose, 10μg/ml adenine and 20 μg/ml uracil. Host strains transformed by vectors comprising the above-described ADH2 promoter may be grown for expression in a rich medium consisting of 1% yeast extract, 2% peptone, and 1% glucose supplemented with 80 μg/ml adenine and 80 μg/ml uracil. Derepression of the ADH2 promoter occurs upon exhaustion of medium glucose. Crude yeast supernatants are harvested by filtration and held at 4°C prior to further purification.

Various mammalian or insect cell culture systems can be employed to express recombinant protein. Baculovirus systems for production of heterologous proteins in insect cells are reviewed by Luckow and Summers, *Bio/Technology 6*:47 (1988). Established cell lines of mammalian origin may be employed. Examples of suitable mammalian host cell lines include the COS-7 line of monkey kidney cells (ATCC CRL 1651), described by Gluzman (*Cell 23*:175, 1981), CV-1 cells (ATCC CCL 70) also derived from monkey kidney, L cells, C127, 3T3, Chinese hamster ovary (CHO), HeLa and BHK cell lines. Mammalian expression vectors may comprise non-transcribed elements such as an origin of replication, a suitable promoter and enhancer linked to the gene to be expressed, and other 5' or 3' flanking nontranscribed sequences, and 5' or 3' nontranslated sequences, such as necessary ribosome binding sites, a poly-adenylation site, splice donor and acceptor sites, and transcriptional termination sequences.

The transcriptional and translational control sequences in expression vectors to be used in transforming vertebrate cells may be provided by viral sources. For example, commonly used promoters and enhancers are derived from Polyoma, Adenovirus 2, Simian Virus 40 (SV40), and human cytomegalovirus. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early and late promoter, enhancer, splice, and polyadenylation sites may be used to provide the other genetic elements required for expression of a heterologous DNA sequence. The early and late promoters are particularly useful because both are obtained easily from the virus as a fragment which also contains the SV40 viral origin or replication (Fiers et al.,

Nature 273:113, 1978). Smaller or larger SV40 fragments may also be used, provided the approximately 250 bp sequence extending from the Hind III site toward the *BglI* site located in the viral origin of replication is included. Exemplary vectors can be constructed as disclosed by Okayama and Berg (*Mol. Cell. Biol. 3*:280, 1983). A useful system for stable high level expression of mammalian receptor cDNAs in C127 murine mammary epithelial cells can be constructed substantially as described by Cosman et al. (*Mol. Immunol. 23*:935, 1986). Other expression vectors for use in mammalian host cells are derived from retroviruses.

Producing and Purifying the *ork* Protein

The present invention provides substantially homogeneous *ork* protein, which may be produced by recombinant expression systems as described above or purified from naturally occurring cells. The ork protein is purified to substantial homogeneity, as indicated by a single protein band upon analysis by SDS-polyacrylamide gel electrophoresis (SDS-PAGE).

In one embodiment of the present invention, *ork* is purified from a cellular source using any suitable protein purification technique. The tissues identified in example 3 as containing *ork* RNA (preferably placental or lung tissue) from a mammalian species of interest may be employed as sources of *ork*, for example.

An alternative process for producing the recombinant *ork* protein of the present invention comprises culturing a host cell transformed with an expression vector comprising a DNA sequence that encodes said *ork* protein under conditions that promote expression of the *ork* protein, which is then purified from culture media or cell extracts. Any suitable purification process may be employed, with the procedure of choice varying according to such factors as the type of host cells and whether or not the desired protein is secreted from the host cells. The fusion protein will be secreted into the culture medium when it is initially fused to a signal sequence or leader peptide operative in the host cells, or when the protein comprises soluble forms of the *ork* polypeptides.

For example, supernatants from expression systems which secrete recombinant protein into the culture medium can be first concentrated using a commercially available protein concentration filter, e.g., an Amicon or Millipore Pellicon ultrafiltration unit. Following the concentration step, the concentrate can be applied to a suitable purification matrix. For example, an immunoaffinity column comprising antibodies directed against *ork* and bound to a suitable support may be employed. A monoclonal antibody specific for *ork* may be prepared as described in example 4. Alternatively, an anion exchange resin can be employed, for example, a matrix or substrate having pendant diethylaminoethyl (DEAE) groups. The matrices can be acrylamide, agarose,

dextran, cellulose or other types commonly employed in protein purification. Alternatively, a cation exchange step can be employed. Suitable cation exchangers include various insoluble matrices comprising sulfopropyl or carboxymethyl groups. Sulfopropyl groups are preferred. One or more reversed-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, e.g., silica gel having pendant methyl or other aliphatic groups, can be employed to further purify *ork*.

Recombinant protein produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more concentration, salting-out, aqueous ion exchange or size exclusion chromatography steps. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of recombinant fusion proteins can disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Fermentation of yeast which express *ork* as a secreted protein greatly simplifies purification. Secreted recombinant protein resulting from a large-scale fermentation can be purified by methods analogous to those disclosed by Urdal et al. (*J. Chromatog.* *296*:171, 1984), involving two sequential, reversed-phase HPLC steps for purification of a recombinant protein on a preparative HPLC column.

Some or all of the foregoing purification steps, in various combinations, can be employed to provide an essentially homogeneous recombinant protein. Recombinant cell culture enables the production of the *ork* protein free of those contaminating proteins which may be normally associated with *ork* as it is found in nature, e.g., in cells, cell exudates or body fluids. The foregoing purification procedures are among those that may be employed to purify non-recombinant *ork* proteins of the present invention as well.

Variants and Derivatives of *ork*

Variants and derivatives of native *ork* proteins that retain the desired biological activity are also within the scope of the present invention. An *ork* variant, as referred to herein, is a polypeptide substantially homologous to a native *ork*, but which has an amino acid sequence different from that of native *ork* (human, murine or other mammalian species) because of one or a plurality of deletions, insertions or substitutions.

The variant amino acid sequence preferably is at least 80% identical to a native *ork* amino acid sequence, most preferably at least 90% identical. When the variant *ork* protein comprises extracellular, transmembrane and cytoplasmic domains, these percent identities apply to the entire sequence and also to the extracellular domain when taken

alone. Since the cytoplasmic domains are relatively conserved among members of the RTK family, the percent identity of the extracellular domain is important in identifying a variant as an *ork* protein.

To illustrate this point, a comparison of the full length *ork* and *tie* amino acid sequences (aligned in Figure 5) reveals 76% identity for the cytoplasmic domains, whereas the percent identity drops to 47.5% for the full length sequences as a whole. The percent identity drops to 33% when just the extracellular domains are compared. A DNA probe corresponding to the extracellular region of *tie* would not hybridize to *ork* DNA under moderately stringent hybridization conditions.

The degree of homology (percent identity) may be determined, for example, by comparing sequence information using the GAP computer program, version 6.0 described by Devereux et al. (*Nucl. Acids Res.* 12:387, 1984) and available from the University of Wisconsin Genetics Computer Group (UWGCG). The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443, 1970), as revised by Smith and Waterman (*Adv. Appl. Math* 2:482, 1981). Briefly, the GAP program defines similarity as the number of aligned symbols (i.e., nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. The preferred default parameters for the GAP program include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) for nucleotides, and the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res.* 14:6745, 1986, as described by Schwartz and Dayhoff, eds., *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, pp. 353-358, 1979; (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3) no penalty for end gaps.

Alterations of the native amino acid sequence may be accomplished by any of a number of known techniques. Mutations can be introduced at particular loci by synthesizing oligonucleotides containing a mutant sequence, flanked by restriction sites enabling ligation to fragments of the native sequence. Following ligation, the resulting reconstructed sequence encodes an analog having the desired amino acid insertion, substitution, or deletion.

Alternatively, oligonucleotide-directed site-specific mutagenesis procedures can be employed to provide an altered gene having particular codons altered according to the substitution, deletion, or insertion required. Exemplary methods of making the alterations set forth above are disclosed by Walder et al. (*Gene* 42:133, 1986); Bauer et al. (*Gene* 37:73, 1985); Craik (*BioTechniques*, January 1985, 12-19); Smith et al. (*Genetic Engineering: Principles and Methods*, Plenum Press, 1981); and U.S. Patent Nos. 4,518,584 and 4,737,462, which are incorporated by reference herein.

Variants may comprise conservatively substituted sequences, meaning that a given amino acid residue is replaced by a residue having similar physiochemical characteristics. Examples of conservative substitutions include substitution of one aliphatic residue for another, such as Ile, Val, Leu, or Ala for one another, or substitutions of one polar residue for another, such as between Lys and Arg; Glu and Asp; or Gln and Asn. Other such conservative substitutions, for example, substitutions of entire regions having similar hydrophobicity characteristics, are well known.

ork also may be modified to create *ork* derivatives by forming covalent or aggregative conjugates with other chemical moieties, such as glycosyl groups, lipids, phosphate, acetyl groups and the like. Covalent derivatives of *ork* may be prepared by linking the chemical moieties to functional groups on *ork* amino acid side chains or at the N-terminus or C-terminus of an *ork* polypeptide or the extracellular domain thereof. Other derivatives of *ork* within the scope of this invention include covalent or aggregative conjugates of *ork* or its fragments with other proteins or polypeptides, such as by synthesis in recombinant culture as N-terminal or C-terminal fusions. For example, the conjugate may comprise a heterologous signal or leader polypeptide sequence at the N-terminus of an *ork* polypeptide. Examples of such signal peptides are the α-factor leader of *Saccharomyces*; the signal sequence for interleukin-7 (IL-7) described in United States Patent 4,965,195; or the signal sequence for interleukin-2 receptor described in United States Patent Application 06/626,667 filed on July 2, 1984. The signal or leader peptide co-translationally or post-translationally directs transfer of the conjugate from its site of synthesis to a site outside of the cell membrane or cell wall.

ork polypeptide fusions can comprise peptides added to facilitate purification and identification of *ork* . Such peptides include, for example, poly-His or the antigenic identification peptides described in U.S. Patent No. 5,011,912 and in Hopp et al., *Bio/Technology* 6:1204, 1988. One such peptide is the FLAG® peptide, Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys (DYKDDDDK), which is highly antigenic and provides an epitope reversibly bound by a specific monoclonal antibody enabling rapid assay and facile purification of expressed recombinant protein. This sequence is also specifically cleaved by bovine mucosal enterokinase at the residue immediately following the Asp-Lys pairing. Fusion proteins capped with this peptide may also be resistant to intracellular degradation in *E. coli*. A murine hybridoma designated 4E11 produces a monoclonal antibody that binds the peptide DYKDDDDK in the presence of certain divalent metal cations (as described in U.S. Patent 5,011,912) and has been deposited with the American Type Culture Collection under accession no HB 9259.

The present invention further includes *ork* polypeptides with or without associated native-pattern glycosylation. *ork* expressed in yeast or mammalian

expression systems (e.g., COS-7 cells) may be similar to or significantly different from a native *ork* polypeptide in molecular weight and glycosylation pattern, depending upon the choice of expression system. Expression of *ork* polypeptides in bacterial expression systems, such as *E. coli*, provides non-glycosylated molecules.

DNA constructs that encode various additions or substitutions of amino acid residues or sequences, or deletions of terminal or internal residues or sequences not needed for biological activity or binding, can be prepared. For example, N-glycosylation sites in the *ork* extracellular domain can be modified to preclude glycosylation while allowing expression of a homogeneous, reduced carbohydrate analog using yeast expression systems. N-glycosylation sites in eukaryotic polypeptides are characterized by an amino acid triplet Asn-X-Y, wherein X is any amino acid except Pro and Y is Ser or Thr. Appropriate modifications to the nucleotide sequence encoding this triplet will result in substitutions, additions or deletions that prevent attachment of carbohydrate residues at the Asn side chain. Known procedures for inactivating N-glycosylation sites in proteins include those described in U.S. Patent 5,071,972 and EP 276,846. In another example, sequences encoding Cys residues that are not essential for biological activity can be altered to cause the Cys residues to be deleted or replaced with other amino acids, preventing formation of incorrect intramolecular disulfide bridges upon renaturation. Other variants are prepared by modification of adjacent dibasic amino acid residues to enhance expression in yeast systems in which KEX2 protease activity is present. EP 212,914 discloses the use of site-specific mutagenesis to inactivate KEX2 protease processing sites in a protein.

Naturally occurring *ork* variants are also encompassed by the present invention. Examples of such variants are proteins that result from alternative mRNA splicing events (since *ork* is encoded by a multi-exon gene) or from proteolytic cleavage of the *ork* protein, provided the desired biological activity (e.g., binding to an anti-*ork* antibody or to the ligand) is retained. Alternative splicing of mRNA may yield a truncated but biologically active *ork* protein, such as a naturally occurring soluble form of the protein, for example. An alternative splicing event in *tie*, an RTK of the same subgroup as *ork*, yielded a protein lacking the first of three fibronectin type III repeats. Variations attributable to proteolysis include, for example, differences in the N- or C-termini upon expression in different types of host cells, due to proteolytic removal of one or more terminal amino acids (which may occur intracellularly or during purification). Varying N-termini may also result from cleavage of the signal peptide in certain host cells at a point other than between amino acids 18 and 19 of the disclosed sequence.

In certain host cells, post-translational processing will remove the methionine residue encoded by an initiation codon, whereas the methionine residue will remain at

the N-terminus of proteins produced in other host cells. The N-terminal amino acid may, for example, be any of the amino acids at positions 1 to 5 of SEQ ID NO:1 (for proteins comprising a signal peptide) or 19-23 (for the mature protein). The C-terminus may be truncated deliberately during expression vector construction (e.g., in constructing vectors encoding soluble proteins as described above) or as a result of differential processing which may remove up to about five C-terminal amino acids, for example.

Due to the known degeneracy of the genetic code wherein more than one codon can encode the same amino acid, a DNA sequence may vary from that presented in SEQ ID NO:1, and still encode an *ork* protein having the amino acid sequence of SEQ ID NO:1. Such variant DNA sequences may result from silent mutations (e.g., occurring during PCR amplification), and may be the product of deliberate mutagenesis of a native sequence.

Nucleic acid sequences within the scope of the present invention include isolated DNA and RNA sequences that hybridize to the *ork* nucleotide sequences disclosed herein under conditions of moderate or severe stringency, and which encode biologically active *ork*. Moderate stringency hybridization conditions refer to conditions described in, for example, Sambrook et al. *Molecular Cloning: A Laboratory Manual,* 2 ed. Vol. 1, pp. 1.101-104, Cold Spring Harbor Laboratory Press, (1989). Conditions of moderate stringency, as defined by Sambrook et al., include use of a prewashing solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0) and hybridization conditions of about 55°C, 5 X SSC, overnight. Conditions of severe stringency include higher temperatures of hybridization and washing. The skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as the length of the probe.

The present invention thus provides isolated DNA sequences encoding biologically active *ork* , selected from: (a) DNA derived from the coding region of a native mammalian *ork* gene (e.g., cDNA derived from the coding region of the human *ork* cDNA sequence presented in SEQ ID NO:1); (b) DNA capable of hybridizing under moderately stringent conditions to a DNA derived from the extracellular region of the sequence presented in SEQ ID NO:1, and which encodes biologically active *ork* ; and (c) DNA which is degenerate as a result of the genetic code to a DNA defined in (a) or (b) and which encodes biologically active *ork*.

ork polypeptides in the form of oligomers such as dimers or trimers are within the scope of the present invention. Oligomers may be linked by disulfide bonds formed between cysteine residues on different *ork* polypeptides. In one embodiment of the invention, an *ork* dimer is created by fusing *ork* to the Fc region of an antibody (IgG1). The Fc polypeptide preferably is fused to the C-terminus of a soluble *ork* (comprising

only the extracellular domain). A gene fusion encoding the *ork* fusion protein is inserted into an appropriate expression vector. The *ork* fusion proteins are allowed to assenble much like antibody molecules, whereupon interchain disulfide bonds form between Fc polypeptides, yielding divalent *ork*. If fusion proteins are made with both heavy and light chains of an antibody, it is possible to form an *ork* oligomer with as many as four *ork* extracellular regions. Alternatively, one can link two soluble ork domains with a peptide linker such as the $Gly_4SerGly_5Ser$ linker sequence described in United States Patent 5,073,627. A fusion protein comprising two or more *ork* polypeptides (with or without peptide spacers) may be produced by recombinant DNA technology.

The present invention further provides antisense or sense oligonucleotides comprising a single-stranded nucleic acid sequence (either RNA or DNA) capable of binding to target *ork* mRNA (sense) or *ork* DNA (antisense) sequences. Antisense or sense oligonucleotides, according to the present invention, comprise a fragment of the coding region of *ork* cDNA. Such a fragment generally comprises at least about 14 nucleotides, preferably from about 14 to about 30 nucleotides. The ability to create an antisense or a sense oligonucleotide, based upon a cDNA sequence for a given protein is described in, for example, Stein and Cohen, *Cancer Res.* 48:2659, 1988 and van der Krol et al., *BioTechniques* 6:958, 1988.

Binding of antisense or sense oligonucleotides to target nucleic acid sequences results in the formation of duplexes that block translation (RNA) or transcription (DNA) by one of several means, including enhanced degradation of the duplexes, premature termination of transcription or translation, or by other means. The antisense oligonucleotides thus may be used to block expression of *ork* proteins. Antisense or sense oligonucleotides further comprise oligonucleotides having modified sugar-phosphodiester backbones (or other sugar linkages, such as those described in WO91/06629) and wherein such sugar linkages are resistant to endogenous nucleases. Such oligonucleotides with resistant sugar linkages are stable *in vivo* (i.e., capable of resisting enzymatic degradation) but retain sequence specificity to be able to bind to target nucleotide sequences. Other examples of sense or antisense oligonucleotides include those oligonucleotides which are covalently linked to organic moieties, such as those described in WO 90/10448, and other moieties that increase affinity of the oligonucleotide for a target nucleic acid sequence, such as poly-(L-lysine). Further still, intercalating agents, such as ellipticine, and alkylating agents or metal complexes may be attached to sense or antisense oligonucleotides to modify binding specificities of the antisense or sense oliginucleotide for the target nucleotide sequence.

Antisense or sense oligonucleotides may be introduced into a cell containing the target nucleic acid sequence by any gene transfer method, including, for example,

$CaPO_4$-mediated DNA transfection, electroporation, or other gene transfer vectors such as Epstein-Barr virus. Antisense or sense oligonucleotides are preferably introduced into a cell containing the target nucleic acid sequence by insertion of the antisense or sense oligonucleotide into a suitable retroviral vector, then contacting the cell with the retrovirus vector containing the inserted sequence, either *in vivo* or *ex vivo*. Suitable retroviral vectors include, but are not limited to, the murine retrovirus M-MuLV, N2 (a retrovirus derived from M-MuLV), or or the double copy vectors designated DCT5A, DCT5B and DCT5C (see PCT Application US 90/02656). Alternatively, other promotor sequences may be used to express the oligonucleotide.

Sense or antisense oligonucleotides may also be introduced into a cell containing the target nucleotide sequence by formation of a conjugate with a ligand binding molecule, as described in WO 91/04753. Suitable ligand binding molecules include, but are not limited to, cell surface receptors, growth factors, other cytokines, or other ligands that bind to cell surface receptors.

Alternatively, a sense or an antisense oligonucleotide may be introduced into a cell containing the target nucleic acid sequence by formation of an oligonucleotide-lipid complex, as described in WO 90/10448. The sense or antisense oligonucleotide-lipid complex is preferably dissociated within the cell by an endogenous lipase.

Uses of Ork Proteins

One use of *ork* is as a research tool for identifying the ligand that binds thereto and studying the biological effects of ligand binding. The *ork* polypeptides of the present invention also may be employed in *in vitro* assays for detection of *ork* or its ligand or the interactions thereof.

The *ork* polypeptides of the present invention can be used in a binding assay to detect cells expressing a ligand for *ork*. For example, *ork* or an extracellular domain or a fragment thereof can be conjugated to a detectable moiety such as ^{125}I. Radiolabeling with ^{125}I can be performed by any of several standard methodologies that yield a functional ^{125}I *ork* molecule labeled to high specific activity. Alternatively, another detectable moiety such as an enzyme that can catalyze a colorometric or fluorometric reaction, biotin or avidin may be used. Cells to be tested for *ork* ligand expression can be contacted with the labeled *ork*. After incubation, unbound labeled *ork* is removed and binding is measured using the detectable moiety.

Soluble *ork* polypeptides may be employed to competitively bind the ligand *in vivo*, thus inhibiting signal transduction activity via endogenous cell surface bound *ork* proteins. Further, soluble proteins are generally more suitable for intravenous administration and may exert their desired effect (e.g. binding a ligand) in the bloodstream. Soluble *ork* proteins comprising only the extracellular (ligand-binding)

domain lack the tyrosine kinase domain which is located within the cytoplasmic domain.

The *ork* polypeptides disclosed herein are also useful in generating antibodies specific to the *ork* immunogens. Monoclonal antibodies specific for *ork* or antigenic fragments thereof may be prepared by procedures that include those described in Example 4. The monoclonal antibodies can be attached to insoluble support materials for use in immunoaffinity column purification of *ork* proteins. Other uses for the antibodies include detecting *ork* proteins in *in vitro* assays and identifying and purifying additional *ork* polypeptides such as variants (e.g., from alternative splicing events) that comprise the region from which the immunogen was derived.

EXAMPLE 1: PCR-Based Cloning of Human *ork* cDNA

Degenerate oligonucleotide primers were synthesized corresponding to the conserved sequences HRDLAA (TK-1; sense orientation) and SDVWS (TK-2; antisense orientation) contained within the kinase domain of all RTKs (Hanks, *et al.*, 1988). The oligonucleotides corresponding to the conserved motifs HRDLAA and SDVWS were
(5'-CAC$^{C}/_{A}$AG$^{G}/_{A}$GA$^{C}/_{T}$$^{C}/_{T}$TGGC$^{A}/_{T}$GC-3') and
(5'-A$^{G}/_{A}$GACCA$^{A/T}/_{C/G}$AC$^{G}/_{A}$TC$^{G}/_{A}$CT-3'), respectively. Both oligonucleotides were 32-fold degenerate and contained recognition sequences for *Xho* I. The HRDLAA sequence was chosen to favor RTKs over *src*-family tyrosine kinases that have the sequence HRDLRA (Hanks *et al.*, 1988).

Single-stranded cDNA was synthesized from human placental polyadenylated RNA by standard methods. The single stranded cDNA was used as template for a PCR reaction using the conditions of Wilks (1989). The two degenerate oligonucleotides were used as primers in the PCR reaction. Amplified PCR reaction products of about 200 base pairs were digested with *Xho* I and ligated into an *Xho* I-digested plasmid vector desinated pBLUESCRIPT® SK. This vector, available from Stratagene Cloning Systems, La Jolla, California, is replicable in *E. coli* and contains a polylinker segment that includes 21 unique restriction sites, one of which is *Xho* I. The ligation mixture was transformed into *E. coli* cells by conventional procedures.

Recombinant plasmids were recovered from the transformed *E. coli* cells and the nucleotide sequence of the DNA insert in a number of individual clones was determined. Among the known RTKs that were represented in the PCR-library were c-*fms*, JTK-4 (Partanen *et al. Proc. Natl. Acad. Sci. USA*, 87:8913-17, 1990), *kdr* (Terman *et al., Oncogene*, 6:1677-1683, 1991), *eph* (Hirai *et al.*, 1987 *supra*), and *flt*-4 (Aprelikova *et al. Cancer Res.*, 52:746-748, 1992). However, one clone, designated HPK-6, contained a novel sequence. The approximately 200bp cDNA

insert of HPK-6 was isolated and radiolabeled with ^{32}P by conventional techniques for use as a probe to isolate longer cDNA sequences encoding the novel protein.

An oligo-dT-primed human placental cDNA library in plasmid pDC302 has been previously described (Larsen *et al., J. Exp. Med.*, 172:1559-1570, 1990). Briefly, total cell RNA was isolated from whole fresh placental tissue and polyadenylated RNA was prepared by chromatography on oligo(dT)-cellulose. Double-stranded, oligo(dT)-primed cDNA was prepared with a commercial kit (Amersham Corp., Arlington Heights, IL). The resulting cDNA was size fractionated by chromatography on Sephacryl S-1000 (Pharmacia Fine Chemicals, Piscataway, NJ) in 0.5 M sodium acetate. The excluded cDNA was cloned into the *BglII* site of the mammalian expresion vector, pDC302 (described below) by an adaptor method similar to that described by Haymerle et al. (*Nucl. Acids Res.* 14:8615, 1986). Briefly, noncomplementary oligonucleotides of the sequence 5'-GATCTTGGAACGAGACGACCTGCT and 5'-AGCAGGTCGTCTCGTTCCAA synthesized on a DNA synthesizer (model 380A; Applied Biosystems, Foster City, CA) were annealed and ligated in separate reactions to either cDNA or *BglII* cut vector. Nonligated oligonucleotides were separated from cDNA or vector by chromatography over Sepharose CL-2B (Pharmacia Fine Chemicals) at 65°C in 10mM Tris (pH 8.0), 0.1mM EDTA. 5 ng of adaptored vector was ligated to adaptored cDNA in 10-µl reactions containing 50 mM sodium chloride, 50 mM Tris-HCl (pH 7.5), 10 mM magnesium chloride, 1 mM spermidine, 0.5 mM ATP, 0.1 U/µl T4 polynucleotide kinase and 0.4 U/µl T4 DNA ligase for 30 min at 37°C. Reactions were then desalted by drop dialysis on VSWP 013 filters (Millipore Corp., Bedford, MA) against distilled water for 40 min immediately before electroporation into *Escherichia coli* strain DH5α. Transformants were obtained with an average cDNA insert size of 1.6 kb.

The pDC302 expression vector employed in preparing this cDNA library has been described by Mosley et al. (*Cell* 59:335, 1989). pDC302 is an expression vector for use in mammalian host cells, but also replicates in *E. coli*.

pDC302 was assembled from pDC201 (Sims et al., *Science* 241:585, 1988), SV40 and cytomegalovirus DNA and comprises, in order with the direction of transcription from the origin of replication: (1) SV40 sequences from coordinates 5171-270 including the origin of replication, enhancer sequences and early and late promoters; (2) human cytomegalovirus sequences including the promoter and enhancer regions (nucleotides 671 to 63 from the sequence published by Boechart et al. (*Cell* 41:521, 1985); (3) adenovirus-2 sequences containing the first exon and part of the intron between the first and second exons of the tripartite leader, the second exon and part of the third exon of the tripartite leader; and (4) a multiple cloning site (MCS) containing sites for XhoI, Asp718, SmaI, NotI and BglII; (5) SV40 sequences from

coordinates 4127-4100 and 2770-2533 that include the polyadenylation and termination signals for early transcription; (6) sequences derived from pBR322 and virus-associated sequences VAI and VAII of pDC201, with adenovirus sequences 10532-11156 containing the VAI and VAII genes, followed by pBR322 sequences from 4363-2486 and 1094-375 containing the ampicillin resistance gene and origin of replication.

Miniprep DNA isolated from individual pools of colonies from the above-described placental cDNA library was digested with *Bgl* II (which excises the cDNA inserts) and screened for hybridization to the HPK-6-derived probe. A clone comprising a 2.3 Kb cDNA insert was isolated from one pool (#398) and the insert was radiolabeled for use in subsequent library screens.

A random-primed cDNA library in λgt10 was generated from poly(A)$^+$ human placental RNA and screened for hybridization to the 2.3 kb cDNA probe. One clone (32-1), about 4.0 Kb in length, contained an entire coding region, as well as 5' and 3' untranslated sequences.

A DNA sequence for human *ork* is presented in SEQ ID NO:1, along with the amino acid sequence encoded thereby. The SEQ ID NO:1 sequence was derived by combining DNA sequencing information from the fully-sequenced 2.3kb clone with that obtained for the 4.0 kb clone (for which certain portions overlapping with the 2.3 kb clone were not fully sequenced). Beginning with an ATG codon at nucleotides 149-151 there is an open reading frame extending for 1124 amino acids. Several lines of evidence suggest that this methionine codon is the initiating codon. The ATG codon is in a proper context (Kozak, 1984), and is followed by a hydrophobic stretch of amino acids that resembles a leader sequence (amino acids 1-18 of SEQ ID NO:1, i.e., methionine through glycine). There are two in-frame termination codons upstream of the ATG codon. The protein comprises an N-terminal extracellular (ligand-binding) domain (amino acids 19-745), followed by a transmembrane region comprising amino acids 746-772, and a C-terminal cytoplasmic domain (which contains the tyrosine kinase domain responsible for the tyrosine phosphorylating activity of the protein) comprising amino acids 773-1124.

The 2.3 kb clone encodes a C-terminal fragment of *ork* , including a small C-terminal portion of the extracellular domain, followed by the complete transmembrane and cytoplasmic domains of the protein (amino acids 698-1124). The 2.3 kb *ork* cDNA extends from nucleotide 2240 to the 3' end of the SEQ ID NO:1 DNA sequence and contains additional 3' non-coding sequences (i.e., the 2.3 kb clone has a longer 3' non-coding region than does the 4.0 kb clone depicted in SEQ ID NO:1). Since the 2.3 kb clone lacks most of the extracellular domain, the tyrosine kinase encoded thereby is not expected to bind a ligand.

The 4.0 kb *ork* cDNA in plasmid pBLUESCRIPT®SK in *E. coli* strain DH5α was deposited with the American Type Culture collection on May 28, 1992 and was given accession number ATCC 69003. The strain deposit was made under the terms of the Budapest Treaty.

EXAMPLE 2: Characterization of *ork* Gene Product

Rabbit polyclonal antiserum was generated against a peptide consisting of the carboxyl-terminal 21 amino acids of the *ork* protein cytoplasmic domain (SEQ ID NO:1 sequence) conjugated to ovalbumin (Kitagawa and Aikawa, *J. Biochem.*79:233-236, 1976). Animals were immunized with 100 μg of the conjugate in complete Freund's adjuvant, followed by three boosts in incomplete adjuvant, after which time test bleeds were taken and assayed for immunoreactivity towards the peptide immobilized on nitrocellulose. Bleeds that showed anti-peptide activity were used in the following immunoprecipitation experiment.

To examine the protein encoded by the *ork* cDNA, the full length cDNA was excised from λgt10 by *Not* I digestion and ligated into a *Not* I-digested expression plasmid pDC302 (described above). *Not* I cleaves pDC302 at a unique site within a multiple cloning site. The resulting recombinant vectors were transfected into COS-7 cells and 3 days later the cells were labelled with ^{35}S-met/cys and detergent solubilized. The ^{35}S-labeled lysates were incubated with 5 μl of the anti-peptide serum for 1 hour at 4°C, after which time 100 μl of pre-swollen proteinA-sepharose was added and the lysates were incubated for an additional hour at 4°C. The immune complexes were washed 3 times with PBS/1% Triton X-100, boiled in SDS-PAGE sample buffer, and analyzed by SDS/polyacrylamide gel electrophoresis. The gel was treated with Amplify (Amersham, Arlington Heights, IL) and exposed to X-Ray film.

As shown in Figure 3 (panel A), the immune serum, but not the pre-immune serum, precipitated an approximately 140 Kd protein in the lysates from the *ork*-transfected cells, but not in lysates from control cells. Preincubation of the immune serum with the peptide specifically blocked the precipitation of the 140 Kd protein.

A hallmark of protein tyrosine kinases is their ability to auto-phosphorylate *in vitro* in immune complexes (Ullrich and Schlessinger, 1990)). We tested the ability of the *ork* protein to autophosphorylate using the rabbit serum described above. *Ork*- or mock-transfected COS-7 cells were washed 1X in cold PBS/0.1mM sodium orthovanadate and lysed at 10^7 cells/ml in a buffer containing: 25mM TRIS pH 8/150mM NaCl/1mM EGTA/1mM DTT/1% NP-40/0.1mM sodium orthovanadate/1mM PMSF/10 μg/ml leupeptin/10 μg/ml pepstatin A. Lysates were incubated with 5 μl of the above-described rabbit anti-peptide serum or normal rabbit

serum for 2 hours at 4°C with 15 µl (packed volume) protein A-sepharose. Immune complexes were washed 3X with 25mM TRIS pH 8/150mM NaCl/1% Triton X-100/0.1% SDS/1% sodium deoxycholate/1mM DTT/0.1mM sodium orthovanadate/1mM PMSF/ 10 µg/ml leupeptin, followed by three washes in 20mM HEPES pH 7.4/10mM $MnCl_2$/5mM $MgCl_2$. *In vitro* phosphorylation reactions were initiated by suspending the beads to a final volume of 50 µl in a mixture containing [γ-^{32}P]ATP (25mM)/10mM $MnCl_2$/5mM $MgCl_2$/20mM HEPES pH 7.4. After incubation for 30 minutes at 37°C the beads were washed with 20mM HEPES pH 7.4/10mM ATP/5mM EGTA, boiled in SDS-PAGE sample buffer and separated on 8-16% SDS/polyacrylamide gels followed by autoradiography. For phosphoaminoacid analysis, separated proteins were electrophoretically transferred to membranes and regions containing ^{32}P-labelled protein were excised and incubated with 6N HCl for 1 hour at 110°C. Hydrolysates were analysed by two dimensional thin-layer electrophoresis (Cooper *et al.*, 1983) and autoradiography.

As shown in Figure 3, panel B, immunoprecipitated *ork* was capable of auto-phosphorylation. Phospho-amino acid analysis demonstrated that only phospho-tyrosine was present in the immunoprecipitates (Fig. 3, panel C). We also examined whether the *ork* expressed in the COS cells was phosphorylated on tyrosine residues. Lysates from *ork*- and control-transfected cells were immunoprecipated with the P1 (polyclonal) serum, separated by SDS-PAGE, transferred to filters and incubated with an anti-phosphotyrosine anti-serum. As shown in Figure 3, panel D, a 140 Kd protein containing phosphotyrosine was seen only in the lysates from *ork*-transfected cells that had been immunoprecipitated with the anti-*ork* serum. It is unclear from these data whether *ork* has a high level of intrinsic tyrosine kinase activity, whether the COS cells express a potential ligand for *ork* which is interacting with the *ork* protein on the COS cell surface, or whether *ork* is being phosphorylated by a tyrosine kinase expressed by COS cells.

EXAMPLE 3: Screening Human Tissues for *ork* Expression

RNAs from several human tissues were screened for *ork* transcripts. For *ork* expression, a filter containing 2µg of poly(A)$^+$ RNA from various human tissues (Clontech, San Diego, CA) was hybridized with an antisense RNA probe containing nucleotides 2240-3005 (outside of the kinase domain). As shown in Figure 4, a 4.5 kb mRNA was found at highest levels in placenta and lung, with lower levels detected in kidney and heart. Each of these tissues is highly vascularized, suggesting that *ork* may be expressed in endothelial cells. Consistent with this, umbilical cord vein endothelial cells (HUVECs) also expressed *ork* message. In addition to the tissues

listed in Figure 4, RNAs isolated from hematopoietic cell lines and tissues were surveyed, but *ork* mRNA was not detected.

EXAMPLE 4: Monoclonal Antibodies Directed Against *ork*

This example illustrates the preparation of monoclonal antibodies to *ork* or immunogenic fragments thereof. Purified *ork* can be used to generate anti-*ork* monoclonal antibodies using conventional techniques, for example, those techniques described in U.S. Patent 4,411,993.

In one embodiment of the invention, the peptide used to generate rabbit polyclonal antiserum in example 2 is used to produce a monoclonal antibody. The peptide comprises the carboxy-terminal 21 amino acids of the *ork* protein (i.e., the last 21 amino acids of the cytoplasmic domain in the SEQ ID NO:1 sequence) conjugated to ovalbumin. Briefly, mice are immunized with the immunogenic peptide emulsified in complete Freund's adjuvant, and injected in amounts ranging from 10-100 μg subcutaneously or intraperitoneally. Ten to twelve days later, the immunized animals are boosted with additional immunogenic peptide emulsified in incomplete Freund's adjuvant. Mice are periodically boosted thereafter on a weekly to bi-weekly immunization schedule. Serum samples are periodically taken by retro-orbital bleeding or tail-tip excision for testing by dot blot assay or ELISA (Enzyme-Linked Immunosorbent Assay), for anti-*ork* antibodies.

Following detection of an appropriate antibody titer, positive animals are provided one last intravenous injection of the immunogen in saline. Three to four days later, the animals are sacrificed, spleen cells harvested, and spleen cells are fused to a murine myeloma cell line (e.g., NS1 or Ag 8.653). Fusions generate hybridoma cells, which are plated in multiple microtiter plates in a HAT (hypoxanthine, aminopterin and thymidine) selective medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

The hybridoma cells are screened by ELISA for reactivity against purified *ork* peptide by adaptations of the techniques disclosed in Engvall et al., *Immunochem.* 8:871, 1971 and in U.S. Patent 4,703,004. Positive hybridoma cells can be injected intraperitoneally into syngeneic BALB/c mice to produce ascites containing high concentrations of anti-*ork* monoclonal antibodies. Alternatively, hybridoma cells can be grown *in vitro* in flasks or roller bottles by various techniques. Monoclonal antibodies produced in mouse ascites can be purified by ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively, affinity chromatography based upon binding of antibody to protein A or protein G can also be used, as can affinity chromatography based upon binding to *ork*.

EXAMPLE 5: Cloning of Murine *ork* cDNA

Murine *ork* cDNA was isolated by cross-species hybridization using human *ork* cDNA as a probe. A murine cDNA library was prepared from 12 1/2 day old mouse embryos. The library was screened by conventional techniques using the radiolabeled 2.3 Kb human *ork* cDNA described in example 1 as a probe.

The DNA sequence of the cDNA insert of a hybridizing clone was determined, and the amino acid sequence encoded thereby is presented in SEQ ID NO:3 and in Figure 6. The murine *ork* sequence (designated "m"; in the top rows) is aligned with the corresponding portion of human *ork* ("h") in Figure 6. Identical amino acids are indicated with a line; conservative or similar amino acid changes are indicated with two dots or one dot, respectively. The percent similarity of these murine and human *ork* amino acid sequences is 99.2% and the percent identity is 97.5%, as determined by the GAP computer program.

The murine *ork* sequence corresponds to amino acids 886-1124 of human *ork*, a region falling within the cytoplasmic domain and extending to the C-terminus of the human *ork* protein. The murine cDNA may be used as a probe to screen the same or a different murine cDNA library to identify a full length murine *ork* clone. Alternatively, given the similarity of the human and murine sequences, murine libraries may be screened using a probe derived from the extracellular region of human *ork*.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Ziegler, Steven F.

(ii) TITLE OF INVENTION: NOVEL TYROSINE KINASE

(iii) NUMBER OF SEQUENCES: 3

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Immunex Corporation
(B) STREET: 51 University Street
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: US
(F) ZIP: 98101

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/905,600
(B) FILING DATE: 26-JUN-1992
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Seese, Kathryn A.
(B) REGISTRATION NUMBER: 32,172
(C) REFERENCE/DOCKET NUMBER: 2609

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (206) 587-0430
(B) TELEFAX: (206) 233-0644
(C) TELEX: 756822

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4138 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 149..3523

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTTCTGTGCT GTTCCTTCTT GCCTCTAACT TGTAAACAAG ACGTACTAGG ACGATGCTAA 60

TGGAAAGTCA CAAACCGCTG GGTTTTTGAA AGGATCCTTG GGACCTCATG CACATTTGTG 120

```
GAAACTGGAT GGAGAGATTT GGGGAAGC ATG GAC TCT TTA GCC AGC TTA GTT      172
                                Met Asp Ser Leu Ala Ser Leu Val
                                  1               5

CTC TGT GGA GTC AGC TTG CTC CTT TCT GGA ACT GTG GAA GGT GCC ATG      220
Leu Cys Gly Val Ser Leu Leu Leu Ser Gly Thr Val Glu Gly Ala Met
     10                  15                  20

GAC TTG ATC TTG ATC AAT TCC CTA CCT CTT GTA TCT GAT GCT GAA ACA      268
Asp Leu Ile Leu Ile Asn Ser Leu Pro Leu Val Ser Asp Ala Glu Thr
 25                  30                  35                  40

TCT CTC ACC TGC ATT GCC TCT GGG TGG CGC CCC CAT GAG CCC ATC ACC      316
Ser Leu Thr Cys Ile Ala Ser Gly Trp Arg Pro His Glu Pro Ile Thr
                 45                  50                  55

ATA GGA AGG GAC TTT GAA GCC TTA ATG AAC CAG CAC CAG GAT CCG CTG      364
Ile Gly Arg Asp Phe Glu Ala Leu Met Asn Gln His Gln Asp Pro Leu
             60                  65                  70

GAA GTT ACT CAA GAT GTG ACC AGA GAA TGG GCT AAA AAA GTT GTT TGG      412
Glu Val Thr Gln Asp Val Thr Arg Glu Trp Ala Lys Lys Val Val Trp
         75                  80                  85

AAG AGA GAA AAG GCT AGT AAG ATC AAT GGT GCT TAT TTC TGT GAA GGG      460
Lys Arg Glu Lys Ala Ser Lys Ile Asn Gly Ala Tyr Phe Cys Glu Gly
     90                  95                  100

CGA GTT CGA GGA GAG GCA ATC AGG ATA CGA ACC ATG AAG ATG CGT CAA      508
Arg Val Arg Gly Glu Ala Ile Arg Ile Arg Thr Met Lys Met Arg Gln
105                 110                 115                 120

CAA GCT TCC TTC CTA CCA GCT ACT TTA ACT ATG ACT GTG GAC AAG GGA      556
Gln Ala Ser Phe Leu Pro Ala Thr Leu Thr Met Thr Val Asp Lys Gly
                125                 130                 135

GAT AAC GTG AAC ATA TCT TTC AAA AAG GTA TTG ATT AAA GAA GAA GAT      604
Asp Asn Val Asn Ile Ser Phe Lys Lys Val Leu Ile Lys Glu Glu Asp
            140                 145                 150

GCA GTG ATT TAC AAA AAT GGT TCC TTC ATC CAT TCA GTG CCC CGG CAT      652
Ala Val Ile Tyr Lys Asn Gly Ser Phe Ile His Ser Val Pro Arg His
        155                 160                 165

GAA GTA CCT GAT ATT CTA GAA GTA CAC CTG CCT CAT GCT CAG CCC CAG      700
Glu Val Pro Asp Ile Leu Glu Val His Leu Pro His Ala Gln Pro Gln
    170                 175                 180

GAT GCT GGA GTG TAC TCG GCC AGG TAT ATA GGA GGA AAC CTC TTC ACC      748
Asp Ala Gly Val Tyr Ser Ala Arg Tyr Ile Gly Gly Asn Leu Phe Thr
185                 190                 195                 200

TCG GCC TTC ACC AGG CTG ATA GTC CGG AGA TGT GAA GCC CAG AAG TGG      796
Ser Ala Phe Thr Arg Leu Ile Val Arg Arg Cys Glu Ala Gln Lys Trp
                205                 210                 215

GGA CCT GAA TGC AAC CAT CTC TGT ACT GCT TGT ATG AAC AAT GGT GTC      844
Gly Pro Glu Cys Asn His Leu Cys Thr Ala Cys Met Asn Asn Gly Val
            220                 225                 230
```

```
TGC CAT GAA GAT ACT GGA GAA TGC ATT TGC CCT CCT GGG TTT ATG GGA      892
Cys His Glu Asp Thr Gly Glu Cys Ile Cys Pro Pro Gly Phe Met Gly
        235                 240                 245

AGG ACG TGT GAG AAG GCT TGT GAA CTG CAC ACG TTT GGC AGA ACT TGT      940
Arg Thr Cys Glu Lys Ala Cys Glu Leu His Thr Phe Gly Arg Thr Cys
    250                 255                 260

AAA GAA AGG TGC AGT GGA CAA GAG GGA TGC AAG TCT TAT GTG TTC TGT      988
Lys Glu Arg Cys Ser Gly Gln Glu Gly Cys Lys Ser Tyr Val Phe Cys
265                 270                 275                 280

CTC CCT GAC CCC TAT GGG TGT TCC TGT GCC ACA GGC TGG AAG GGT CTG     1036
Leu Pro Asp Pro Tyr Gly Cys Ser Cys Ala Thr Gly Trp Lys Gly Leu
                285                 290                 295

CAG TGC AAT GAA GCA TGC CAC CCT GGT TTT TAC GGG CCA GAT TGT AAG     1084
Gln Cys Asn Glu Ala Cys His Pro Gly Phe Tyr Gly Pro Asp Cys Lys
            300                 305                 310

CTT AGG TGC AGC TGC AAC AAT GGG GAG ATG TGT GAT CGC TTC CAA GGA     1132
Leu Arg Cys Ser Cys Asn Asn Gly Glu Met Cys Asp Arg Phe Gln Gly
        315                 320                 325

TGT CTC TGC TCT CCA GGA TGG CAG GGG CTC CAG TGT GAG AGA GAA GGC     1180
Cys Leu Cys Ser Pro Gly Trp Gln Gly Leu Gln Cys Glu Arg Glu Gly
    330                 335                 340

ATA CCG AGG ATG ACC CCA AAG ATA GTG GAT TTG CCA GAT CAT ATA GAA     1228
Ile Pro Arg Met Thr Pro Lys Ile Val Asp Leu Pro Asp His Ile Glu
345                 350                 355                 360

GTA AAC AGT GGT AAA TTT AAT CCC ATT TGC AAA GCT TCT GGC TGG CCG     1276
Val Asn Ser Gly Lys Phe Asn Pro Ile Cys Lys Ala Ser Gly Trp Pro
                365                 370                 375

CTA CCT ACT AAT GAA GAA ATG ACC CTG GTG AAG CCG GAT GGG ACA GTG     1324
Leu Pro Thr Asn Glu Glu Met Thr Leu Val Lys Pro Asp Gly Thr Val
            380                 385                 390

CTC CAT CCA AAA GAC TTT AAC CAT ACG GAT CAT TTC TCA GTA GCC ATA     1372
Leu His Pro Lys Asp Phe Asn His Thr Asp His Phe Ser Val Ala Ile
        395                 400                 405

TTC ACC ATC CAC CGG ATC CTC CCC CCT GAC TCA GGA GTT TGG GTC TGC     1420
Phe Thr Ile His Arg Ile Leu Pro Pro Asp Ser Gly Val Trp Val Cys
    410                 415                 420

AGT GTG AAC ACA GTG GCT GGG ATG GTG GAA AAG CCC TTC AAC ATT TCT     1468
Ser Val Asn Thr Val Ala Gly Met Val Glu Lys Pro Phe Asn Ile Ser
425                 430                 435                 440

GTT AAA GTT CTT CCA AAG CCC CTG AAT GCC CCA AAC GTG ATT GAC ACT     1516
Val Lys Val Leu Pro Lys Pro Leu Asn Ala Pro Asn Val Ile Asp Thr
                445                 450                 455

GGA CAT AAC TTT GCT GTC ATC AAC ATC AGC TCT GAG CCT TAC TTT GGG     1564
Gly His Asn Phe Ala Val Ile Asn Ile Ser Ser Glu Pro Tyr Phe Gly
            460                 465                 470
```

```
GAT GGA CCA ATC AAA TCC AAG AAG CTT CTA TAC AAA CCC GTT AAT CAC      1612
Asp Gly Pro Ile Lys Ser Lys Lys Leu Leu Tyr Lys Pro Val Asn His
        475                 480                 485

TAT GAG GCT TGG CAA CAT ATT CAA GTG ACA AAT GAG ATT GTT ACA CTC      1660
Tyr Glu Ala Trp Gln His Ile Gln Val Thr Asn Glu Ile Val Thr Leu
    490                 495                 500

AAC TAT TTG GAA CCT CGG ACA GAA TAT GAA CTC TGT GTG CAA CTG GTC      1708
Asn Tyr Leu Glu Pro Arg Thr Glu Tyr Glu Leu Cys Val Gln Leu Val
505                 510                 515                 520

CGT CGT GGA GAG GGT GGG GAA GGG CAT CCT GGA CCT GTG AGA CGC TTC      1756
Arg Arg Gly Glu Gly Gly Glu Gly His Pro Gly Pro Val Arg Arg Phe
                525                 530                 535

ACA ACA GCT TCT ATC GGA CTC CCT CCT CCA AGA GGT CTA AAT CTC CTG      1804
Thr Thr Ala Ser Ile Gly Leu Pro Pro Pro Arg Gly Leu Asn Leu Leu
            540                 545                 550

CCT AAA AGT CAG ACC ACT CTA AAT TTG ACC TGG CAA CCA ATA TTT CCA      1852
Pro Lys Ser Gln Thr Thr Leu Asn Leu Thr Trp Gln Pro Ile Phe Pro
        555                 560                 565

AGC TCG GAA GAT GAC TTT TAT GTT GAA GTG GAG AGA AGG TCT GTG CAA      1900
Ser Ser Glu Asp Asp Phe Tyr Val Glu Val Glu Arg Arg Ser Val Gln
    570                 575                 580

AAA AGT GAT CAG CAG AAT ATT AAA GTT CCA GGC AAC TTG ACT TCG GTG      1948
Lys Ser Asp Gln Gln Asn Ile Lys Val Pro Gly Asn Leu Thr Ser Val
585                 590                 595                 600

CTA CTT AAC AAC TTA CAT CCC AGG GAG CAG TAC GTG GTC CGA GCT AGA      1996
Leu Leu Asn Asn Leu His Pro Arg Glu Gln Tyr Val Val Arg Ala Arg
                605                 610                 615

GTC AAC ACC AAG GCC CAG GGG GAA TGG AGT GAA GAT CTC ACT GCT TGG      2044
Val Asn Thr Lys Ala Gln Gly Glu Trp Ser Glu Asp Leu Thr Ala Trp
            620                 625                 630

ACC CTT AGT GAC ATT CTT CCT CCT CAA CCA GAA AAC ATC AAG ATT TCC      2092
Thr Leu Ser Asp Ile Leu Pro Pro Gln Pro Glu Asn Ile Lys Ile Ser
        635                 640                 645

AAC ATT ACA CAC TCC TCG GCT GTG ATT TCT TGG ACA ATA TTG GAT GGC      2140
Asn Ile Thr His Ser Ser Ala Val Ile Ser Trp Thr Ile Leu Asp Gly
    650                 655                 660

TAT TCT ATT TCT TCT ATT ACT ATC CGT TAC AAG GTT CAA GGC AAG AAT      2188
Tyr Ser Ile Ser Ser Ile Thr Ile Arg Tyr Lys Val Gln Gly Lys Asn
665                 670                 675                 680

GAA GAC CAG CAC GTT GAT GTG AAG ATA AAG AAT GCC ACC ATC ATT CAG      2236
Glu Asp Gln His Val Asp Val Lys Ile Lys Asn Ala Thr Ile Ile Gln
                685                 690                 695

TAT CAG CTC AAG GGC CTA GAG CCT GAA ACA GCA TAC CAG GTG GAC ATT      2284
Tyr Gln Leu Lys Gly Leu Glu Pro Glu Thr Ala Tyr Gln Val Asp Ile
            700                 705                 710
```

```
TTT GCA GAG AAC AAC ATA GGG TCA AGC AAC CCA GCC TTT TCT CAT GAA      2332
Phe Ala Glu Asn Asn Ile Gly Ser Ser Asn Pro Ala Phe Ser His Glu
        715                 720                 725

CTG GTG ACC CTC CCA GAA TCT CAA GCA CCA GCG GAC CTC GGA GGG GGG      2380
Leu Val Thr Leu Pro Glu Ser Gln Ala Pro Ala Asp Leu Gly Gly Gly
    730                 735                 740

AAG ATG CTG CTT ATA GCC ATC CTT GGC TCT GCT GGA ATG ACC TGC CTG      2428
Lys Met Leu Leu Ile Ala Ile Leu Gly Ser Ala Gly Met Thr Cys Leu
745                 750                 755                 760

ACT GTG CTG TTG GCC TTT CTG ATC ATA TTG CAA TTG AAG AGG GCA AAT      2476
Thr Val Leu Leu Ala Phe Leu Ile Ile Leu Gln Leu Lys Arg Ala Asn
                765                 770                 775

GTG CAA AGG AGA ATG GCC CAA GCC TTC CAA AAC GTG AGG GAA GAA CCA      2524
Val Gln Arg Arg Met Ala Gln Ala Phe Gln Asn Val Arg Glu Glu Pro
            780                 785                 790

GCT GTG CAG TTC AAC TCA GGG ACT CTG GCC CTA AAC AGG AAG GTC AAA      2572
Ala Val Gln Phe Asn Ser Gly Thr Leu Ala Leu Asn Arg Lys Val Lys
        795                 800                 805

AAC AAC CCA GAT CCT ACA ATT TAT CCA GTG CTT GAC TGG AAT GAC ATC      2620
Asn Asn Pro Asp Pro Thr Ile Tyr Pro Val Leu Asp Trp Asn Asp Ile
    810                 815                 820

AAA TTT CAA GAT GTG ATT GGG GAG GGC AAT TTT GGC CAA GTT CTT AAG      2668
Lys Phe Gln Asp Val Ile Gly Glu Gly Asn Phe Gly Gln Val Leu Lys
825                 830                 835                 840

GCG CGC ATC AAG AAG GAT GGG TTA CGG ATG GAT GCT GCC ATC AAA AGA      2716
Ala Arg Ile Lys Lys Asp Gly Leu Arg Met Asp Ala Ala Ile Lys Arg
                845                 850                 855

ATG AAA GAA TAT GCC TCC AAA GAT GAT CAC AGG GAC TTT GCA GGA GAA      2764
Met Lys Glu Tyr Ala Ser Lys Asp Asp His Arg Asp Phe Ala Gly Glu
            860                 865                 870

CTG GAA GTT CTT TGT AAA CTT GGA CAC CAT CCA AAC ATC ATC AAT CTC      2812
Leu Glu Val Leu Cys Lys Leu Gly His His Pro Asn Ile Ile Asn Leu
        875                 880                 885

TTA GGA GCA TGT GAA CAT CGA GGC TAC TTG TAC CTG GCC ATT GAG TAC      2860
Leu Gly Ala Cys Glu His Arg Gly Tyr Leu Tyr Leu Ala Ile Glu Tyr
    890                 895                 900

GCG CCC CAT GGA AAC CTT CTG GAC TTC CTT CGC AAG AGC CGT GTG CTG      2908
Ala Pro His Gly Asn Leu Leu Asp Phe Leu Arg Lys Ser Arg Val Leu
905                 910                 915                 920

GAG ACG GAC CCA GCA TTT GCC ATT GCC AAT AGC ACC GCG TCC ACA CTG      2956
Glu Thr Asp Pro Ala Phe Ala Ile Ala Asn Ser Thr Ala Ser Thr Leu
                925                 930                 935

TCC TCC CAG CAG CTC CTT CAC TTC GCT GCC GAC GTG GCC CGG GGC ATG      3004
Ser Ser Gln Gln Leu Leu His Phe Ala Ala Asp Val Ala Arg Gly Met
            940                 945                 950
```

```
GAC TAC TTG AGC CAA AAA CAG TTT ATC CAC AGG GAT CTG GCT GCC AGA      3052
Asp Tyr Leu Ser Gln Lys Gln Phe Ile His Arg Asp Leu Ala Ala Arg
        955                 960                 965

AAC ATT TTA GTT GGT GAA AAC TAT GTG GCA AAA ATA GCA GAT TTT GGA      3100
Asn Ile Leu Val Gly Glu Asn Tyr Val Ala Lys Ile Ala Asp Phe Gly
    970                 975                 980

TTG TCC CGA GGT CAA GAG GTG TAC GTG AAA AAG ACA ATG GGA AGG CTC      3148
Leu Ser Arg Gly Gln Glu Val Tyr Val Lys Lys Thr Met Gly Arg Leu
985                 990                 995                 1000

CCA GTG CGC TGG ATG GCC ATC GAG TCA CTG AAT TAC AGT GTG TAC ACA      3196
Pro Val Arg Trp Met Ala Ile Glu Ser Leu Asn Tyr Ser Val Tyr Thr
                1005                1010                1015

ACC AAC AGT GAT GTA TGG TCC TAT GGT GTG TTA CTA TGG GAG ATT GTT      3244
Thr Asn Ser Asp Val Trp Ser Tyr Gly Val Leu Leu Trp Glu Ile Val
            1020                1025                1030

AGC TTA GGA GGC ACA CCC TAC TGC GGG ATG ACT TGT GCA GAA CTC TAC      3292
Ser Leu Gly Gly Thr Pro Tyr Cys Gly Met Thr Cys Ala Glu Leu Tyr
        1035                1040                1045

GAG AAG CTG CCC CAG GGC TAC AGA CTG GAG AAG CCC CTG AAC TGT GAT      3340
Glu Lys Leu Pro Gln Gly Tyr Arg Leu Glu Lys Pro Leu Asn Cys Asp
    1050                1055                1060

GAT GAG GTG TAT GAT CTA ATG AGA CAA TGC TGG CGG GAG AAG CCT TAT      3388
Asp Glu Val Tyr Asp Leu Met Arg Gln Cys Trp Arg Glu Lys Pro Tyr
1065                1070                1075                1080

GAG AGG CCA TCA TTT GCC CAG ATA TTG GTG TCC TTA AAC AGA ATG TTA      3436
Glu Arg Pro Ser Phe Ala Gln Ile Leu Val Ser Leu Asn Arg Met Leu
                1085                1090                1095

GAG GAG CGA AAG ACC TAC GTG AAT ACC ACG CTT TAT GAG AAG TTT ACT      3484
Glu Glu Arg Lys Thr Tyr Val Asn Thr Thr Leu Tyr Glu Lys Phe Thr
            1100                1105                1110

TAT GCA GGA ATT GAC TGT TCT GCT GAA GAA GCG GCC TAGGACAGAA           3530
Tyr Ala Gly Ile Asp Cys Ser Ala Glu Glu Ala Ala
        1115                1120                112

CATCTGTATA CCCTCTGTTT CCCTTTCACT GGCATGGGAG ACCCTTGACA ACTGCTGAGA   3590

AAACATGCCT CTGCCAAAGG ATGTGATATA TAAGTGTACA TATGTGCTGG AATTCTAACA   3650

AGTCATAGGT TAATATTTAA GACACTGAAA AATCTAAGTG ATATAAATCA GATTCTTCTC   3710

TCTCATTTTA TCCCTCACCT GTAGCATGCC AGTCCCGTTT CATTTAGTCA TGTGACCACT   3770

CTGTCTTGTG TTTCCACAGC CTGCAAGTTC AGTCCAGGAT GCTAACATCT AAAAATAGAC   3830

TTAAATCTCA TTGCTTACAA GCCTAAGAAT CTTTAGAGAA GTATACATAA GTTTAGGATA   3890

AAATAATGGG ATTTTCTTTT CTTTTCTCTG GTAATATTGA CTTGTATATT TTAAGAAATA   3950

ACAGAAAGCC TGGGTGACAT TTGGGAGACA TGTGACATTT ATATATTGAA TTAATATCCC   4010

TACATGTATT GCACATTGTA AAAAGTTTTA GTTTTGATGA GTTGTGAGTT TACCTTGTAT   4070
```

ACTGTAGGCA CACTTTGCAC TGATATATCA TGAGTGAATA AATGTCTTGC CTACTCAAAA 4130

AAAAAAAA 4138

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1124 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asp Ser Leu Ala Ser Leu Val Leu Cys Gly Val Ser Leu Leu Leu
1 5 10 15

Ser Gly Thr Val Glu Gly Ala Met Asp Leu Ile Leu Ile Asn Ser Leu
20 25 30

Pro Leu Val Ser Asp Ala Glu Thr Ser Leu Thr Cys Ile Ala Ser Gly
35 40 45

Trp Arg Pro His Glu Pro Ile Thr Ile Gly Arg Asp Phe Glu Ala Leu
50 55 60

Met Asn Gln His Gln Asp Pro Leu Glu Val Thr Gln Asp Val Thr Arg
65 70 75 80

Glu Trp Ala Lys Lys Val Val Trp Lys Arg Glu Lys Ala Ser Lys Ile
85 90 95

Asn Gly Ala Tyr Phe Cys Glu Gly Arg Val Arg Gly Glu Ala Ile Arg
100 105 110

Ile Arg Thr Met Lys Met Arg Gln Gln Ala Ser Phe Leu Pro Ala Thr
115 120 125

Leu Thr Met Thr Val Asp Lys Gly Asp Asn Val Asn Ile Ser Phe Lys
130 135 140

Lys Val Leu Ile Lys Glu Glu Asp Ala Val Ile Tyr Lys Asn Gly Ser
145 150 155 160

Phe Ile His Ser Val Pro Arg His Glu Val Pro Asp Ile Leu Glu Val
165 170 175

His Leu Pro His Ala Gln Pro Gln Asp Ala Gly Val Tyr Ser Ala Arg
180 185 190

Tyr Ile Gly Gly Asn Leu Phe Thr Ser Ala Phe Thr Arg Leu Ile Val
195 200 205

Arg Arg Cys Glu Ala Gln Lys Trp Gly Pro Glu Cys Asn His Leu Cys
210 215 220

Thr Ala Cys Met Asn Asn Gly Val Cys His Glu Asp Thr Gly Glu Cys
225 230 235 240

Ile Cys Pro Pro Gly Phe Met Gly Arg Thr Cys Glu Lys Ala Cys Glu
245 250 255

Leu His Thr Phe Gly Arg Thr Cys Lys Glu Arg Cys Ser Gly Gln Glu
260 265 270

Gly Cys Lys Ser Tyr Val Phe Cys Leu Pro Asp Pro Tyr Gly Cys Ser
275 280 285

Cys Ala Thr Gly Trp Lys Gly Leu Gln Cys Asn Glu Ala Cys His Pro
290 295 300

Gly Phe Tyr Gly Pro Asp Cys Lys Leu Arg Cys Ser Cys Asn Asn Gly
305 310 315 320

Glu Met Cys Asp Arg Phe Gln Gly Cys Leu Cys Ser Pro Gly Trp Gln
325 330 335

Gly Leu Gln Cys Glu Arg Glu Gly Ile Pro Arg Met Thr Pro Lys Ile
340 345 350

Val Asp Leu Pro Asp His Ile Glu Val Asn Ser Gly Lys Phe Asn Pro
355 360 365

Ile Cys Lys Ala Ser Gly Trp Pro Leu Pro Thr Asn Glu Glu Met Thr
370 375 380

Leu Val Lys Pro Asp Gly Thr Val Leu His Pro Lys Asp Phe Asn His
385 390 395 400

Thr Asp His Phe Ser Val Ala Ile Phe Thr Ile His Arg Ile Leu Pro
405 410 415

Pro Asp Ser Gly Val Trp Val Cys Ser Val Asn Thr Val Ala Gly Met
420 425 430

Val Glu Lys Pro Phe Asn Ile Ser Val Lys Val Leu Pro Lys Pro Leu
435 440 445

Asn Ala Pro Asn Val Ile Asp Thr Gly His Asn Phe Ala Val Ile Asn
450 455 460

Ile Ser Ser Glu Pro Tyr Phe Gly Asp Gly Pro Ile Lys Ser Lys Lys
465 470 475 480

Leu Leu Tyr Lys Pro Val Asn His Tyr Glu Ala Trp Gln His Ile Gln
485 490 495

Val Thr Asn Glu Ile Val Thr Leu Asn Tyr Leu Glu Pro Arg Thr Glu
500 505 510

Tyr Glu Leu Cys Val Gln Leu Val Arg Arg Gly Glu Gly Gly Glu Gly
515 520 525

His Pro Gly Pro Val Arg Arg Phe Thr Thr Ala Ser Ile Gly Leu Pro
530 535 540

Pro Pro Arg Gly Leu Asn Leu Leu Pro Lys Ser Gln Thr Thr Leu Asn
545 550 555 560

Leu Thr Trp Gln Pro Ile Phe Pro Ser Ser Glu Asp Asp Phe Tyr Val
565 570 575

Glu Val Glu Arg Arg Ser Val Gln Lys Ser Asp Gln Gln Asn Ile Lys
580 585 590

Val Pro Gly Asn Leu Thr Ser Val Leu Leu Asn Asn Leu His Pro Arg
595 600 605

Glu Gln Tyr Val Val Arg Ala Arg Val Asn Thr Lys Ala Gln Gly Glu
610 615 620

Trp Ser Glu Asp Leu Thr Ala Trp Thr Leu Ser Asp Ile Leu Pro Pro
625 630 635 640

Gln Pro Glu Asn Ile Lys Ile Ser Asn Ile Thr His Ser Ser Ala Val
645 650 655

Ile Ser Trp Thr Ile Leu Asp Gly Tyr Ser Ile Ser Ser Ile Thr Ile
660 665 670

Arg Tyr Lys Val Gln Gly Lys Asn Glu Asp Gln His Val Asp Val Lys
675 680 685

Ile Lys Asn Ala Thr Ile Ile Gln Tyr Gln Leu Lys Gly Leu Glu Pro
690 695 700

Glu Thr Ala Tyr Gln Val Asp Ile Phe Ala Glu Asn Asn Ile Gly Ser
705 710 715 720

Ser Asn Pro Ala Phe Ser His Glu Leu Val Thr Leu Pro Glu Ser Gln
725 730 735

Ala Pro Ala Asp Leu Gly Gly Gly Lys Met Leu Leu Ile Ala Ile Leu
740 745 750

Gly Ser Ala Gly Met Thr Cys Leu Thr Val Leu Leu Ala Phe Leu Ile
755 760 765

Ile Leu Gln Leu Lys Arg Ala Asn Val Gln Arg Arg Met Ala Gln Ala
770 775 780

Phe Gln Asn Val Arg Glu Glu Pro Ala Val Gln Phe Asn Ser Gly Thr
785 790 795 800

Leu Ala Leu Asn Arg Lys Val Lys Asn Asn Pro Asp Pro Thr Ile Tyr
805 810 815

Pro Val Leu Asp Trp Asn Asp Ile Lys Phe Gln Asp Val Ile Gly Glu
820 825 830

Gly Asn Phe Gly Gln Val Leu Lys Ala Arg Ile Lys Lys Asp Gly Leu
835 840 845

Arg Met Asp Ala Ala Ile Lys Arg Met Lys Glu Tyr Ala Ser Lys Asp
850 855 860

Asp His Arg Asp Phe Ala Gly Glu Leu Glu Val Leu Cys Lys Leu Gly
865 870 875 880

```
His His Pro Asn Ile Ile Asn Leu Leu Gly Ala Cys Glu His Arg Gly
                885                 890                 895

Tyr Leu Tyr Leu Ala Ile Glu Tyr Ala Pro His Gly Asn Leu Leu Asp
            900                 905                 910

Phe Leu Arg Lys Ser Arg Val Leu Glu Thr Asp Pro Ala Phe Ala Ile
        915                 920                 925

Ala Asn Ser Thr Ala Ser Thr Leu Ser Ser Gln Gln Leu Leu His Phe
    930                 935                 940

Ala Ala Asp Val Ala Arg Gly Met Asp Tyr Leu Ser Gln Lys Gln Phe
945                 950                 955                 960

Ile His Arg Asp Leu Ala Ala Arg Asn Ile Leu Val Gly Glu Asn Tyr
                965                 970                 975

Val Ala Lys Ile Ala Asp Phe Gly Leu Ser Arg Gly Gln Glu Val Tyr
            980                 985                 990

Val Lys Lys Thr Met Gly Arg Leu Pro Val Arg Trp Met Ala Ile Glu
        995                 1000                1005

Ser Leu Asn Tyr Ser Val Tyr Thr Thr Asn Ser Asp Val Trp Ser Tyr
    1010                1015                1020

Gly Val Leu Leu Trp Glu Ile Val Ser Leu Gly Gly Thr Pro Tyr Cys
1025                1030                1035                1040

Gly Met Thr Cys Ala Glu Leu Tyr Glu Lys Leu Pro Gln Gly Tyr Arg
                1045                1050                1055

Leu Glu Lys Pro Leu Asn Cys Asp Asp Glu Val Tyr Asp Leu Met Arg
            1060                1065                1070

Gln Cys Trp Arg Glu Lys Pro Tyr Glu Arg Pro Ser Phe Ala Gln Ile
        1075                1080                1085

Leu Val Ser Leu Asn Arg Met Leu Glu Glu Arg Lys Thr Tyr Val Asn
    1090                1095                1100

Thr Thr Leu Tyr Glu Lys Phe Thr Tyr Ala Gly Ile Asp Cys Ser Ala
1105                1110                1115                1120

Glu Glu Ala Ala
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 239 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Val Pro Trp Leu Gly Ala Cys Glu His Arg Gly Tyr Leu Tyr Leu Ala
1               5                  10                  15

Ile Glu Tyr Ala Pro His Gly Asn Leu Leu Asp Phe Leu Arg Lys Ser
            20                  25                  30

Arg Val Leu Glu Thr Asp Pro Ala Phe Ala Ile Ala Asn Ser Thr Ala
        35                  40                  45

Ser Ile Met Ser Ser Gln Gln Leu Leu His Phe Ala Ala Asp Val Ala
    50                  55                  60

Arg Gly Met Asp Tyr Leu Ser Gln Lys Gln Phe Ile His Arg Asp Leu
65                  70                  75                  80

 Ala Ala Arg Asn Ile Leu Val Gly Glu Asn Tyr Ile Ala Lys Ile Ala
                85                  90                  95

Asp Phe Gly Leu Ser Arg Gly Gln Glu Val Tyr Val Lys Lys Thr Met
            100                 105                 110

Gly Arg Leu Pro Val Arg Trp Met Ala Ile Glu Ser Leu Asn Tyr Ser
        115                 120                 125

Val Tyr Thr Thr Asn Ser Asp Val Trp Ser Tyr Gly Val Leu Leu Trp
    130                 135                 140

Glu Ile Val Ser Leu Gly Gly Thr Pro Tyr Cys Gly Met Thr Cys Ala
145                 150                 155                 160

Glu Leu Tyr Glu Lys Leu Pro Gln Gly Tyr Arg Leu Glu Lys Pro Leu
                165                 170                 175

Asn Cys Asp Asp Glu Val Tyr Asp Leu Met Arg Gln Cys Trp Arg Glu
            180                 185                 190

Lys Pro Tyr Glu Arg Pro Ser Phe Ala Gln Ile Leu Val Ser Leu Asn
        195                 200                 205

Arg Met Leu Glu Glu Arg Lys Thr Tyr Val Asn Thr Thr Leu Tyr Glu
    210                 215                 220

Lys Phe Thr Tyr Ala Gly Ile Asp Cys Ser Ala Glu Glu Ala Ala
225                 230                 235
```

CLAIMS

What is claimed is:

1. An isolated DNA sequence encoding a biologically active human *ork* polypeptide, wherein the DNA sequence is selected from the group consisting of:

(a) DNA comprising the coding region of the DNA sequence of SEQ ID NO:1;

(b) DNA sequences that hybridize under moderately stringent conditions to the DNA of (a) and which encode a biologically active *ork*; and

(c) DNA sequences that are degenerate as a result of the genetic code to a DNA sequence of (a) or (b), and which encode a biologically active *ork*.

2. An isolated DNA sequence according to claim 1 wherein the *ork* polypeptide is a soluble human *ork* polypeptide.

3. An isolated DNA sequence according to claim 2, wherein said soluble human *ork* polypeptide comprises amino acids 19-745 of the amino acid sequence of SEQ ID NO:1.

4. An isolated DNA sequence comprising a nucleotide sequence selected from the group consisting of nucleotides 149-3520, 203-3520, 149-2383, and 203-2383 of SEQ ID NO:1.

5. An expression vector comprising a DNA sequence according to claim 1.

6. An expression vector comprising a DNA sequence according to claim 2.

7. An expression vector comprising a DNA sequence according to claim 3.

8. An expression vector comprising a DNA sequence according to claim 4.

9. A process for preparing an *ork* polypeptide, comprising culturing a host cell transformed with an expression vector according to claim 5 under conditions promoting expression of *ork*, and recovering the *ork* polypeptide from the culture.

10. A process for preparing an *ork* polypeptide, comprising culturing a host cell transformed with an expression vector according to claim 6 under conditions promoting expression of *ork* and recovering the *ork* polypeptide from the culture.

11. A process for preparing an *ork* polypeptide, comprising culturing a host cell transformed with an expression vectoraccording to claim 7 under conditions promoting expression of *ork*, and recovering the *ork* polypeptide from the culture.

12. A process for preparing an *ork* polypeptide, comprising culturing a host cell transformed with an expression vector according to claim 8 under conditions promoting expression of *ork*, and recovering the *ork* polypeptide from the culture.

13. A substantially homogeneous purified biologically active *ork* protein encoded by a DNA sequence selected from the group consisting of:

(a) DNA comprising the coding region of the DNA sequence of SEQ ID NO:1;

(b) DNA sequences that hybridize under moderately stringent conditions to the DNA of (a) and which encode a biologically active *ork*; and

(c) DNA sequences that are degenerate as a result of the genetic code to a DNA sequence of (a) or (b), and which encode a biologically active *ork*.

14. A purified *ork* according to claim 13, wherein said *ork* is human *ork*.

15. A purified *ork* according to claim 14, wherein said *ork* is a soluble human *ork*.

16. A purified human *ork* protein comprising an amino acid sequence selected from the group consisting of amino acids 19-1124 of SEQ ID NO:2 and amino acids 19-745 of SEQ ID NO:2.

17. A purified *ork* protein comprising an extracellular domain that comprises (from N- to C-terminus) an immunoglobulin-like domain, three EGF-like cysteine repeats, a sequence containing a pair of cysteine residues, and three FNIII-like repeats, wherein the amino acid sequence of said extracellular domain is at least 80% identical to the sequence presented as amino acids 19-745 of SEQ ID NO:1.

18. A purified *ork* protein according to claim 17, wherein said protein additionally contains a transmembrane region and a cytoplasmic domain.

19. An antibody immunoreactive with *ork* or an *ork* immunogen.

20. An antisense or sense oligonucleotide that can inhibit transcription or translation of *ork*, comprising a sequence of at least about 14 nucleotides of the DNA sequence of SEQ ID NO:1, or its DNA or RNA complement.

Figure 1

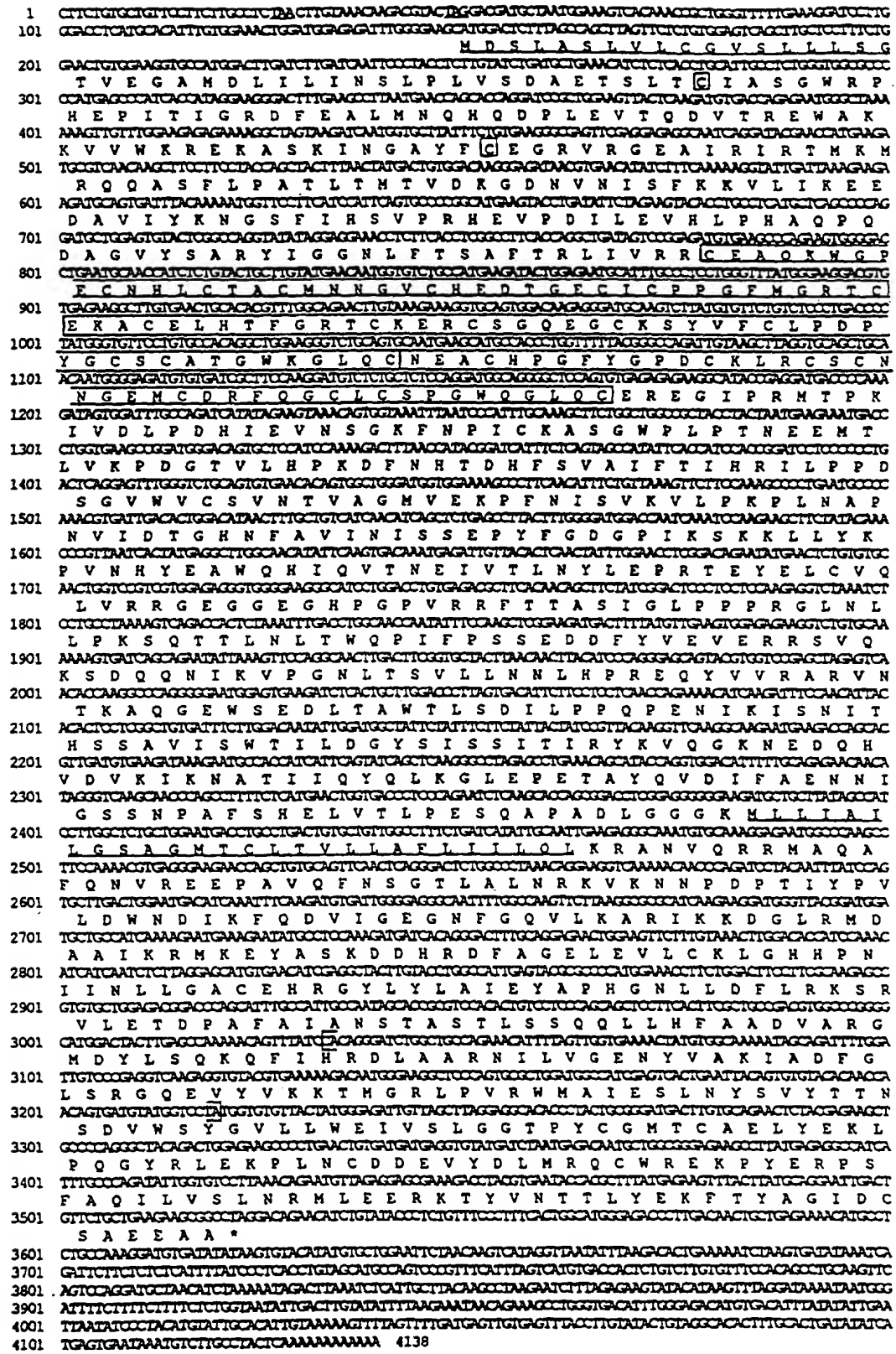


Figure 2

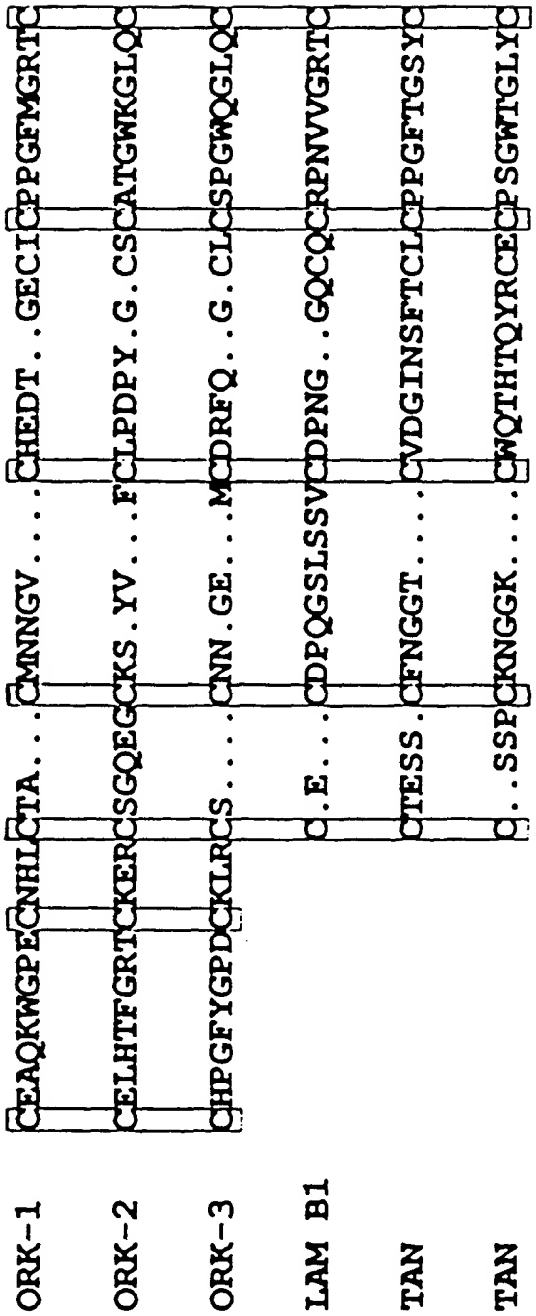


FIGURE 3A

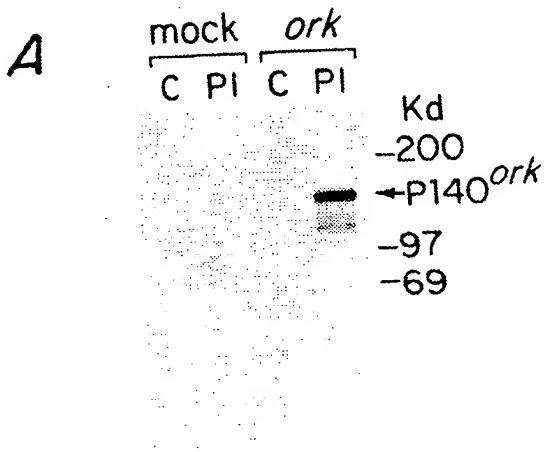


FIGURE 3B

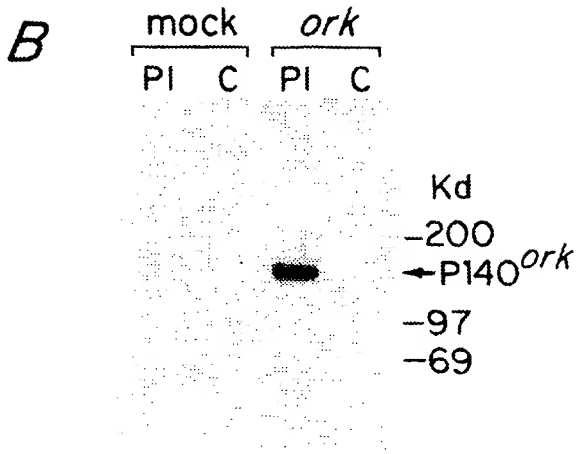


FIGURE 3C

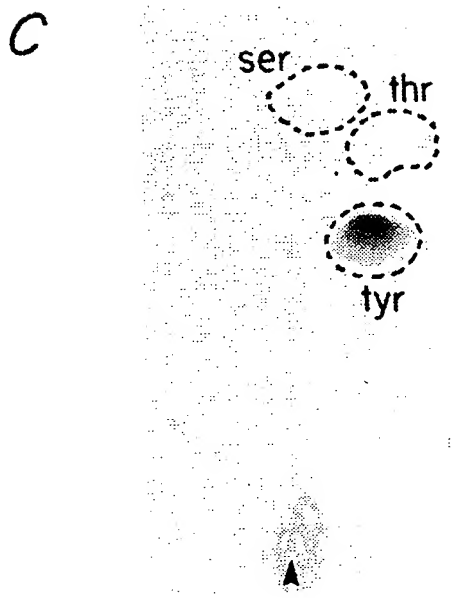


FIGURE 3D

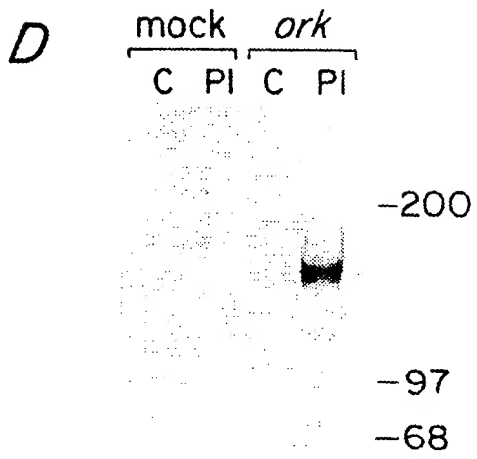


FIGURE 4

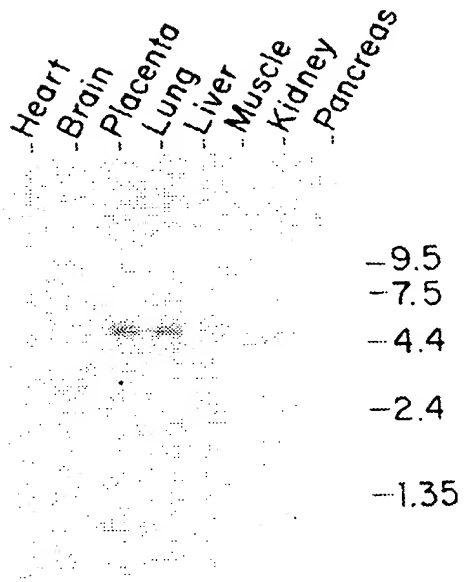
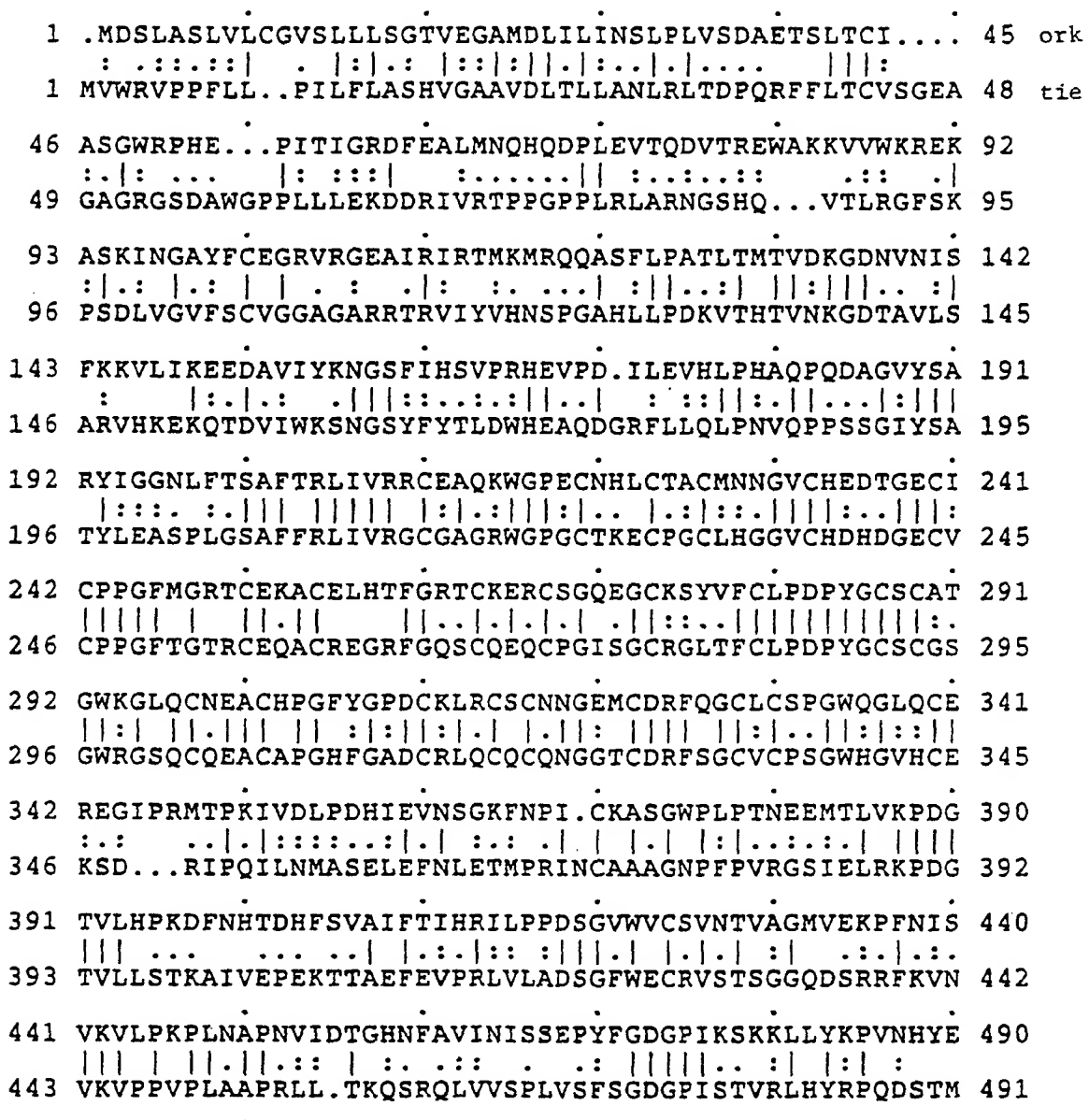


Figure 5a



SUBSTITUTE SHEET

Figure 5b

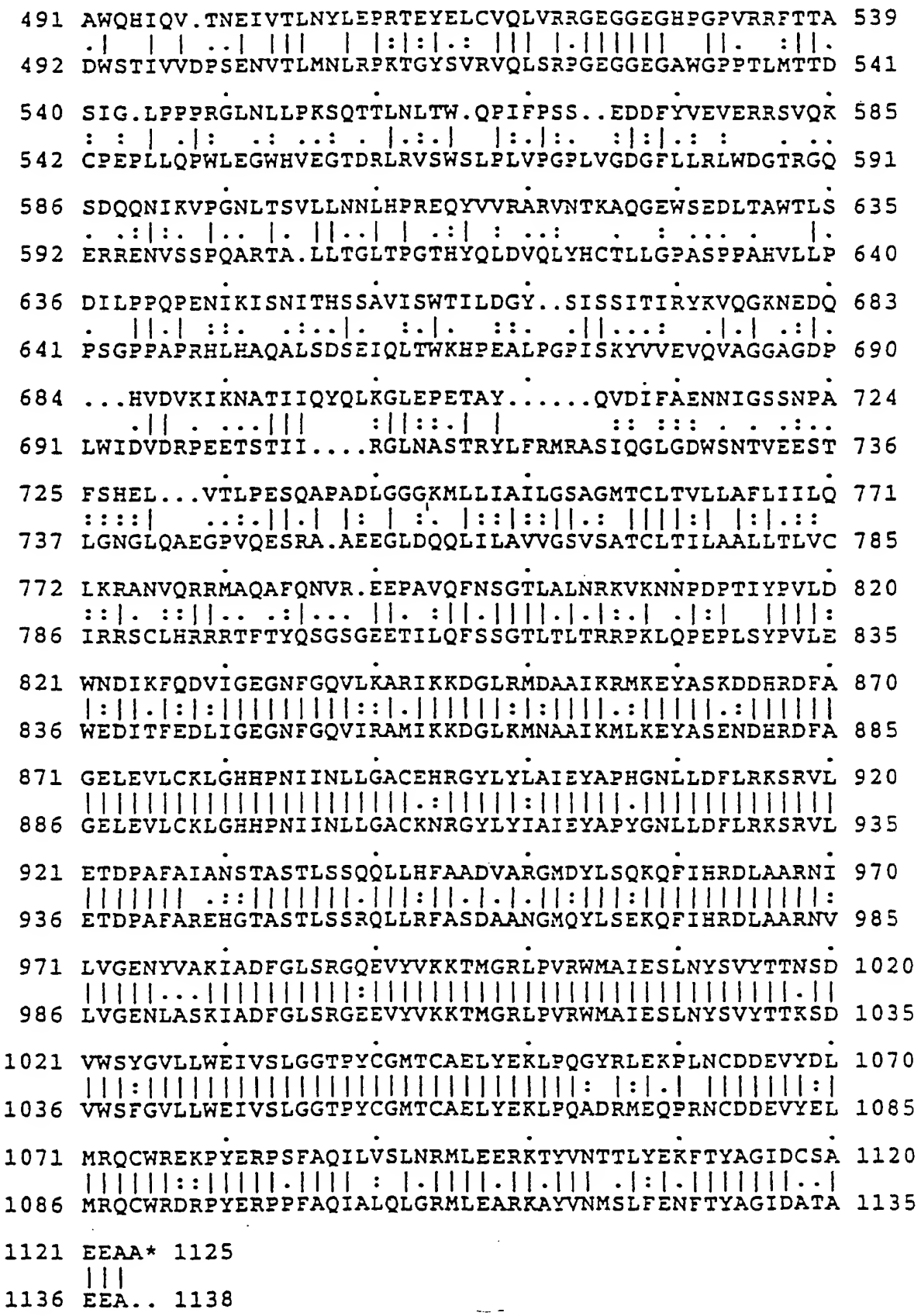


Figure 6

```
  3 VPWLGACEHRGYLYLAIEYAPHGNLLDFLRKSRVLETDPAFAIANSTASI 52  m
    : :|||||||||||||||||||||||||||||||||||||||||||||||.
886 INLLGACEHRGYLYLAIEYAPHGNLLDFLRKSRVLETDPAFAIANSTAST 935 h

 53 MSSQQLLHFAADVARGMDYLSQKQFIHRDLAARNILVGENYIAKIADFGL 102 m
    :|||||||||||||||||||||||||||||||||||||:|||||||||||
936 LSSQQLLHFAADVARGMDYLSQKQFIHRDLAARNILVGENYVAKIADFGL 985 h

103 SRGQEVYVKKTMGRLPVRWMAIESLNYSVYTTNSDVWSYGVLLWEIVSLG 152 m
    ||||||||||||||||||||||||||||||||||||||||||||||||||
986 SRGQEVYVKKTMGRLPVRWMAIESLNYSVYTTNSDVWSYGVLLWEIVSLG 1035 h

153 GTPYCGMTCAELYEKLPQGYRLEKPLNCDDEVYDLMRQCWREKPYERPSF 202 m
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1036 GTPYCGMTCAELYEKLPQGYRLEKPLNCDDEVYDLMRQCWREKPYERPSF 1085 h

203 AQILVSLNRMLEERKTYVNTTLYEKFTYAGIDCSAEEAA 241 m
    |||||||||||||||||||||||||||||||||||||||
1086 AQILVSLNRMLEERKTYVNTTLYEKFTYAGIDCSAEEAA 1124h
```

SUBSTITUTE SHEET

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/06093

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C07H 21/00; C07K 15/00; C12N 9/12, 15/15, 15/54, 15/63; C12P 21/02
US CL :435/69.1, 194, 320.1; 530/387; 536/23.2, 24.5

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 69.2, 194, 320.1; 530/350, 351, 387, 395, 396, 399; 536/23.2, 24.5; 935/9, 13, 14, 34, 36

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	Oncogene, Volume 8, issued 1993, S.F. Ziegler et al., "Molecular Cloning and characterization of a novel receptor protein tyrosine kinase from human placenta", pages 663-670, entire document.	1-20
X	US, A, 4,543,439, (Frackelton et al.) 24 September 1985, entire document.	19
X,P Y	Oncogene, Volume 7, issued 1992, D. L. Dumont et al., "tek, a novel tyrosine kinase gene located on mouse chromosome 4, is expressed in endothelial cells and their presumptive precursors", pages 1471-1480, especially figure 1.	13,17-18 1-12, 14-16, 19-20
Y	Cell, Volume 63, Issued 05 October 1990, J.G. Flanagan et al., "The kit ligand: a cell surface molecule altered in steel mutant fibroblasts", pages 185-194, entire document.	2, 6, 10, 15

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
17 September 1993

Date of mailing of the international search report
29 SEP 1993

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. NOT APPLICABLE

Authorized officer
ROBERT A. WAX

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/06093

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	Proc. Nat. Acad. Sci, U.S.A., Volume 87, issued November 1990, J. Partanen et al., " Putative tyrosine kinases expressed in K-652 human leukemia cells", pages 8913-8917, entire document.	1, 5, 13-14 2, 6, 9-10, 15
X Y	Molecular and Cellular Biology, Volume 12, Number 4, issued April 1992, J. Partanen et al., "A novel endothelial cell surface receptor tyrosine kinase with extracellular epidermal growth factor homology domains, pages 1698-1707, entire document.	1, 5, 9, 13-14 2, 6, 10, 15
Y	Molecular and Cellular Biology, Volume 11, Number 10, issued October 1991, J. P. O'Bryan et al., "axl, a transforming gene isolated from primary human myeloid leukemia cells, encodes a novel receptor tyrosine kinase", pages 5016-5031, entire document.	9-12

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, Dialog ~~files 155, 5, 434~~ (Medline, Biosis, Scisearch), PIR, EMBL, Genbank
search terms: tyrosine kinase??, receptor??, genetics, DNA, orphan, ork, tie, tek, express?, solub?

INTERNATIONAL SEARCH REPORT

Inte...tional application No.
PCT/US93/06093

Box I Observations where certain claims were found unsearchable (Continuation f item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
(Telephone Practice)

I. Claims 1-18, drawn to recombinant DNAs encoding ork, and their expression products, classified in 435/69.1.
II. Claim 19, drawn to an antibody immunoreactive with ork, classified in 530/387.1.
III. Claim 20, drawn to inhibitory oligonucleotides, classified in 536.24.5.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark n Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07H 21/00, C07K 15/00
C12N 9/12, 15/15, 15/54
C12N 15/63, C12P 21/02

A1

(11) International Publication Number: WO 94/00469

(43) International Publication Date: 6 January 1994 (06.01.94)

(21) International Application Number: PCT/US93/06093

(22) International Filing Date: 25 June 1993 (25.06.93)

(30) Priority data:
07/905,600 26 June 1992 (26.06.92) US

(71) Applicant: IMMUNEX CORPORATION [US/US]; 51 University Street, Seattle, WA 98101 (US).

(72) Inventor: ZIEGLER, Steven, F. ; 2230 West Halladay Street, Seattle, WA 98199 (US).

(74) Agent: SEESE, Kathryn, A.; Immunex Corporation, 51 University Street, Seattle, WA 98101 (US).

(81) Designated States: AU, CA, FI, JP, KR, NO, NZ, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: NOVEL TYROSINE KINASE

(57) Abstract

A novel receptor protein tyrosine kinase named *ork* (Orphan receptor tyrosine kinase) is identified and characterized. cDNA encoding the *ork* protein is inserted into an expression vector for production of the protein via recombinant DNA technology. The *ork* cDNA, when transfected into COS-7 cells, encodes a 140Kd protein with *in vitro* kinase activity. The *ork* gene is expressed predominantly in placenta and lung, with lower levels in umbilical vein endothelial cells, brain and kidney.

* (Referred to in PCT Gazette No. 08/1994, Section II)

BNSDOCID: <WO 9400469A1>

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**MR**	Mauritania
AU	Australia	**GA**	Gabon	**MW**	Malawi
BB	Barbados	**GB**	United Kingdom	**NE**	Niger
BE	Belgium	**GN**	Guinea	**NL**	Netherlands
BF	Burkina Faso	**GR**	Greece	**NO**	Norway
BG	Bulgaria	**HU**	Hungary	**NZ**	New Zealand
BJ	Benin	**IE**	Ireland	**PL**	Poland
BR	Brazil	**IT**	Italy	**PT**	Portugal
BY	Belarus	**JP**	Japan	**RO**	Romania
CA	Canada	**KP**	Democratic People's Republic of Korea	**RU**	Russian Federation
CF	Central African Republic			**SD**	Sudan
CG	Congo	**KR**	Republic of Korea	**SE**	Sweden
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovak Republic
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LU**	Luxembourg	**TD**	Chad
CS	Czechoslovakia	**LV**	Latvia	**TG**	Togo
CZ	Czech Republic	**MC**	Monaco	**UA**	Ukraine
DE	Germany	**MG**	Madagascar	**US**	United States of America
DK	Denmark	**ML**	Mali	**UZ**	Uzbekistan
ES	Spain	**MN**	Mongolia	**VN**	Viet Nam
FI	Finland				

NOVEL TYROSINE KINASE

BACKGROUND

The ability of cells to respond to environmental cues is in large part due to the interaction of cell-surface receptors with external stimuli. A number of factors that interact with receptors have been identified, among which are receptor-binding proteins that are soluble, membrane-bound, or exist in both forms.

One class of receptors, the receptor tyrosine kinases (RTKs), has been intensively studied and shown to be crucial to the growth and differentiation of a variety of cell types (Yarden and Ullrich, *Ann. Rev. Biochem.* 57:433-478, 1988). Tyrosine kinases are enzymes that catalyze the phosphorylation of tyrosine residues. Tyrosine phosphorylation is associated with signal-transduction across the cellular plasma membrane. The protein-tyrosine kinase family can be grouped into two very broad families: the above-mentioned RTKs, which are, or are intimately associated with, membrane-spanning growth factor receptors; and those that are associated with the membrane but lack a transmembrane sequence (Yarden and Ullrich, 1988, *supra.*).

The RTKs can be further divided into five subgroups on the basis of structural similarities in their extracellular domains and the organization of the tyrosine kinase catalytic region in their cytoplasmic domains. Subgroups I (epidermal growth factor (EGF) receptor-like), II (insulin receptor-like) and the *eph/eck* family contain cysteine-rich sequences (Hirai *et al., Science* 238:1717-1720, 1987; Yanden and Ullrich, *Ann. Rev. Biochem*, 57,443-478, 1988; Lindberg and Hunter, *Mol. Cell. Biol.* 10:6316-6324, 1990). The functional domains of the kinase region of these three classes of RTKs are encoded as a contiguous sequence (Hanks *et al., Science* 241, 42-52, 1988). Subgroups III (platelet- derived growth factor (PDGF) receptor-like) and IV (the fibroblast growth factor (FGF) receptors) are characterized as having immunoglobulin (Ig)-like folds in their extracellular domains, as well as having their kinase domains divided in two parts by a variable stretch of unrelated amino acids (Yanden and Ullrich, 1988; Hanks *et al.*, 1988).

While all members of the RTK family share a related cytoplasmic catalytic domain, the extracellular, ligand-binding domains of these receptors have adapted patchwork structure utilizing several structural motifs. The variablility in the structure of the ligand-binding domains of the RTKs almost certainly reflects the diversity of the ligands for these receptors (Ullrich and Schlessinger, *Cell* 61, 243-254, 1990). These ligands range from relatively small, soluble peptides to cell surface proteins that

themselves resemble receptors. Examples of ligands bound by certain members of the RTK family are polypeptide growth factors and hormones.

To deal with this diversity of ligands the RTKs have evolved extracellular domains that are a composite of several structural motifs. For example, the extracellular domain of the *axl/ark* gene contains both Ig-domains and fibronectin type III (FNIII) repeats (O'bryan et al., *Mol. Cell. Biol.* 11:5016-5031, 1991; Rescigno et al., *Oncogene* 6:1909-1913, 1991), while members of the *eph* family have those two motifs separated by a (non-EGF-like) cysteine-rich domain (Hirai et al., *Science* 238:1717-1720, 1987; Lindberg et al., *Mol. Cell. Biol.* 10:6316-6324, 1990; Lhotak et al., *Mol. Cell. Biol.* 11:2496-2502, 1991; Chan and Watt, *Oncogene* 6:1057-1061,1991). This diversity strongly suggests that this family of receptors evolved by accumulating the structural motifs needed for ligand binding and combining these motifs with a conserved catalytic domain.

In view of the role tyrosine kinases play in cell growth and differentiation, as well as signal transduction, isolation of novel tyrosine kinases enables one to study such biological processes. Identifying tyrosine kinases and their ligands also permits exploration of methods for inhibiting or enhancing the interaction thereof, depending on the desired biological effect.

SUMMARY OF INVENTION

The present invention provides a novel receptor (a protein tyrosine kinase), isolated DNA encoding the tyrosine kinase, recombinant expression vectors containing the isolated DNA, and host cells transformed with the recombinant vector. Also provided is a method for producing the novel protein by cultivating the transformed host cells under conditions that promote expression of the tyrosine kinase, and recovering the expressed tyrosine kinase.

BRIEF DESCRIPTION OF FIGURES

Figure 1 presents a nucleotide sequence and deduced amino acid sequence of human *ork* cDNA. The nucleotide sequence is derived from two overlapping cDNA clones isolated from placental cDNA libraries. The initiating methionine codon and the transmembrane region are each underlined. The two cysteines of the immunoglobulin domain are boxed, as are the three EGF-like repeats. Brackets enclose the region amplified by PCR using primers based on conserved kinase domain sequences, as described in example 1. Numbering of nucleotides is in the left margin, while numbering of amino acids is on the right. The signal peptide comprises amino acids 1-18, with the threonine residue at position 19 being the first amino acid of the mature protein. The extracellular domain comprises amino acids 19-745, and the cytoplasmic domain comprises amino acids 773-1124.

Figure 2 presents an analysis of the EGF repeats in the *ork* sequence. The three EGF-repeats in the *ork* sequence were aligned with those in the *tan*-1 and laminin B1 genes. The conserved cysteine residues are boxed and other conserved residues (conserved in two *ork* repeats and at least one other repeat) are marked with an asterisk.

Figure 3 presents the results of studies described in example 2 below. The human *ork* cDNA encodes a 140 Kd phosphoprotein with *in vitro* tyrosine kinase activity. **A.** COS cells were transfected with the *ork* cDNA (*ork*), or a control plasmid without cDNA insert (mock), metabolically labelled with ^{35}S-met/cys, and immunoprecipitated with the P1 rabbit antiserum, raised against the carboxy-terminal 21 amino acids of the predicted *ork* amino acid sequence (P1), or with pre-immune serum (C). The position of P140^{ork} is indicated, as are the positions of size standards. **B.** COS cells were transfected and immunoprecipitated as described above. The immunoprecipitates were incubated with ^{32}P-ATP and separated by SDS-PAGE. The positions of P140^{ork} and size standards are indicated. **C.** The gel from Panel B was transferred to a membrane, exposed to X-ray film and the band corresponding to P140^{ork} was excised. The filter slice was boiled in 6N HCl and eluted material was separated by two-dimensional electrophoresis. The positions of standards for phosphoserine, phosphothreonine and phosphotyrosine, as well as the origin, are indicated. **D.** COS cells were transfected and immunoprecipitated as described above. The immunoprecipitates were separated by SDS-PAGE, transferred to a membrane and incubated with an anti-phosphotyrosine antibody. The positions of size standards are indicated.

Figure 4 shows the results of a study of expression of *ork* mRNA in human tissues, as described in example 3 below. Two micrograms of poly(A)-containing RNA from the indicated tissues was hybridized with an ^{32}P-UTP-labelled RNA probe from the *ork* cDNA. The positions of size standards are marked.

Figures 5a and 5b depict a comparison of the amino acid sequences of *ork* (top lines) and *tie* (bottom lines), the latter being a tyrosine kinase of the same subgroup as *ork*. The sequence comparison was generated by the GAP computer program described. Identical amino acids are indicated with a line; conservative or similar amino acid changes are indicated with two dots or one dot, respectively. The percent similarity and percent identity for the *ork* and *tie* sequences were found to be 64.5% and 47.5%, respectively.

Figure 6 presents a comparison of the amino acid sequence encoded by a murine *ork* clone with the corresponding portion of a human *ork* protein.

DETAILED DESCRIPTION OF INVENTION

The present invention provides a novel receptor protein tyrosine kinase, isolated DNA encoding the tyrosine kinase, recombinant expression vectors containing the isolated DNA, and host cells transformed with the recombinant vector. A method for producing the novel protein involves cultivating the transformed host cells under conditions that promote expression of the tyrosine kinase, and recovering the expressed tyrosine kinase from the cell culture.

The novel protein is designated herein as *ork* (*o*rphan *r*eceptor tyrosine *k*inase). An *ork* cDNA isolated from human placenta and transfected into COS-7 cells encodes a 140Kd protein with the following combination of structural motifs in its extracellular domain: an immunoglobulin (Ig)-like domain followed by three epidermal growth factor (EGF)-like cysteine-rich repeats, which in turn are followed by three fibronectin type III (FNIII) repeats. Of the human tissues tested (example 3 below), the *ork* gene is expressed predominately in placenta and lung, with lower levels in umbilical vein endothelial cells, brain and kidney.

Human *ork* is within the scope of the present invention, as are *ork* proteins derived from other mammalian species. As used herein, the term "*ork*" includes membrane-bound proteins (comprising a cytoplasmic domain, a transmembrane region, and an extracellular domain) as well as truncated proteins that retain the desired biological properties. Such truncated proteins include, for example, soluble *ork* comprising only the extracellular (ligand binding) domain.

Using a polymerase chain reaction-based approach we have isolated *ork* cDNA and characterized this novel receptor protein tyrosine kinase. As described in example 1 below, degenerate oligonucleotide probes based on certain sequences that are conserved within the kinase domain of RTKs were used as primers in a PCR reaction. Single-stranded cDNA derived from human placental poly(A)$^+$ RNA was employed as the template. The PCR reaction products were inserted into a cloning vector and the nucleotide sequence of the cDNA inserts in a number of the resulting recombinant vectors was determined. A clone containing a cDNA insert of about 200 basepairs with a novel DNA sequence was identified.

The cDNA insert was excised, radiolabeled, and used to probe an oligo-dT-primed human placental cDNA library. A hybridizing clone comprising a 2.3 Kb insert was identified, sequenced, and found to be a partial clone encoding a C-terminal *ork* fragment. The cDNA insert was excised, radiolabeled, and used to probe a random-primed human placental cDNA library. A clone comprising a cDNA insert about 4.0 Kb in length that contained an entire coding region for the novel protein, as well as 5' and 3' untranslated sequences, was isolated. The 4.0 kb *ork* cDNA in plasmid

pBLUESCRIPT®SK in *E. coli* strain DH5α was deposited with the American Type Culture collection on May 28, 1992 and was given accession number ATCC 69003. The strain deposit was made under the terms of the Budapest Treaty.

A DNA sequence for a human *ork* cDNA is presented in Figure 1, along with the amino acid sequence encoded thereby. Human *ork* DNA and amino acid sequences identical to those of Figure 1 are presented in SEQ ID NO:1 and SEQ ID NO:2. Several features of the *ork* amino acid sequence are noteworthy. The extracellular domain is a patchwork of three structural motifs. Between amino acid residues 211 and 340 are three copies of an EGF-like cysteine repeat (Davis, *New Biologist* 2:410-419, 1991). The EGF-like repeats in the *ork* gene (boxed in Figure 1) differ from the consensus motif in that they have 8 cysteines each instead of 6. In this regard they are closely related to those in the *tie* gene (Partanen et al., *Mol. Cell. Biol.*, 12:1698-1707, 1992). The proteins with the next most closely related cysteine repeats are laminin B1, laminin B2 (Sasaki et al., *J. Biol. Chem* 263:16536-16544, 1988) and *TAN*, the human homolog of the Drosophila *notch* gene (Ellison et al, *Cell* 66:649-661). The three EGF-repeats in the *ork* sequence were aligned with those in the *tan*-1 and laminin B1 genes in Figure 2. In each case the cysteines in the repeat were aligned.

Amino terminal to the EGF repeats are paired cysteines (boxed in Figure 1) indicative of Ig-like domains. This Ig domain is most similar to the C2 type as described by Williams and Barclay, *Ann. Rev. Immunol.* 6:381-405, 1988. A second pair of cysteine residues follow the EGF repeats; however the additional amino acid residues required to generate a proper Ig-fold are missing. At the membrane proximal region of the extracellular domain, between amino acids 440 and 733, are three repeats of the fibronectin type III motif. Thus, the extracellular domain of *ork* is a structural mosaic consisting of repeats of at least three different functional motifs.

Amino acid residues 746-772 are hydrophobic in nature and likely serve as a transmembrane domain. The cytoplasmic domain contains all the amino acid sequence hallmarks of a tyrosine kinase (Hanks et al., *Science* 241:42-52, 1988). A comparison of the *ork* amino acid sequence with known protein tyrosine kinases showed *ork* to be a member of the RTK family.

Within the receptor tyrosine kinase family, the *ork* and above-mentioned *tie* genes form their own subgroup. The *tie* gene (described by Partanen et al., *supra*) has an extracellular domain that, like *ork*, comprises (from N- to C-terminus) an Ig-like domain followed by three EGF-like domains, an incomplete Ig-like domain, and three fibronectin type III repeats next to the transmembrane region. Additionally, a cDNA form lacking the first of the three epidermal growth factor homology domains was isolated by Partanen et al., *supra*, suggesting that alternative splicing creates different

tie-type receptors. COS-7 cells transfected with a *tie* cDNA expression vector produced a glycosylated protein of 117 kDa.

A comparison of the amino acid sequences of *ork* and *tie* is presented in Figures 5a and 5b, which shows them to be distinct proteins. The comparison presented in Figures 5a and 5b was generated using the GAP computer program, which is described in detail below. The percent similarity and percent identity of the *ork* and *tie* amino acid sequences was found to be 64.5% and 47.5%, respectively. The RTK subgroup containing *ork* and *tie* is characterized by a unique arrangement of structural motifs in the proteins' extracellular domain, as described above. There are also several features in the cytoplasmic kinase domain worth noting. There is a short amino acid insert between the two parts of the kinase domains (Fig. 1). This insert is similar in size to that seen in the *ret* gene (Takahashi and Cooper, *Mol. Cell. Biol.* 7:1378-1385, 1987). However, unlike the insert sequences in other RTK subgroups, the insert in the *ork* and *tie* genes contains no tyrosine residues. Several tyrosine residues are found in the cytoplasmic tail. As phosphorylation of tyrosine in either the kinase insert or the cytoplasmic tail is required for association of the RTK molecule with putative signal transduction molecules, this feature of the *ork* and *tie* genes mostly closely resembles the EGFR subfamily where phosphorylated tyrosine residues in the cytoplasmic tail have been shown to associate with other proteins such as phospholipase C-γ (Margolis et al., *Cell* 57:1101-1107, 1989).

One embodiment of the present invention provides an *ork* protein comprising an extracellular domain that comprises (from N- to C-terminus) an immunoglobin-like domain, three EGF-like cysteine-rich repeats, and three FNIII repeats, wherein the amino acid sequence of said extracellular domain is at least 80% identical to the sequence shown as amino acids 19-745 of SEQ ID NO:1. The extracellular domain amino acid sequence preferably is at least 90% identical to the extracellular domain sequence of SEQ ID NO:1. The percent identity between two amino acid sequences may be determined by using the GAP computer program available from the University of Wisconsin and described in detail below. Such *ork* proteins include those having a transmembrane region and a cytoplasmic domain (or a portion thereof) in addition to the extracellular domain. Such proteins also include those in which a sequence containing a pair of cysteines is positioned between the last EGF-like repeat and the first FN III repeat. Such a sequence (a stretch of amino acids resembling an incomplete Ig-like domain) is found in the native human *ork* protein described above. Since this sequence is not believed to function as an Ig-like domain, it is possible that *ork* proteins lacking some or all of this partial Ig-like domain (amino acids 341-439) will retain the ability to bind a ligand. The *ork* proteins when initially synthesized may comprise a signal peptide as well.

The expression pattern of the *ork* gene suggests that it is predominately expressed in endothelial cells. The structure of the extracellular domain, with the three EGF repeats, is especially intriguing in light of this. Several cell adhesion molecules, including endothelial-leukocyte adhesion molecule-1 (ELAM-1) contain EGF repeats. ELAM-1 is expressed on the surface of activated endothelial cells and is involved in the attachment of neutrophils at sites of inflammation (Bevilacqua et al., *Science* 243:1150-1165, 1989; Siegelman et al., *Cell* 61:611-622, 1990). Several adhesion molecules also contain Ig and FNIII repeats as well. This suggests a possible role for the *ork* gene product in the communication between endothelial cells and leukocytes at sites of inflammation. cDNA encoding an *ork* polypeptide may be isolated from other mammalian species by procedures analogous to those employed in isolating the human *ork* clone. For example, a cDNA library derived from another mammalian species may be substituted for the human cDNA library that was screened using the degenerate probes in example 1 below. Alternatively (and preferably), the human *ork* cDNAs isolated in example 1 are labeled and used as probes to screen mammalian cDNA or genomic libraries using cross-species hybridization techniques. The probe may be derived from the coding region of the above-described 2.3kb or 4.0 kb human *ork* cDNAs.

Murine *ork* cDNA was identified by cross-species hybridization. A murine cDNA library was screened using the 2.3 Kb human *ork* cDNA as a probe, as described in example 5.

Cell types from which cDNA and genomic libraries may be prepared include those in which *ork* RNA expression was detected in example 3. mRNAs isolated from various cell lines can be screened by Northern hybridization to determine additional suitable sources of mammalian *ork* mRNA for use in cloning an *ork* gene. Nucleic acid from mammalian sources that include but are not limited to murine, bovine, porcine, and primate, may be screened to identify *ork* genes.

In addition to the membrane-bound full length protein depicted in SEQ ID NO:1, the present invention provides soluble forms of the *ork* protein. "Soluble *ork*" as used in the context of the present invention refers to polypeptides that contain all or part of the extracellular region of a native *ork* and that, due to the absence of a transmembrane region that would cause retention of the polypeptide on a cell membrane, are secreted upon expression. Fragments of the extracellular domain may be employed as long as the fragment possesses the desired biological activity (e.g., binding to an anti-*ork* antibody or to the ligand for *ork*). Soluble *ork* may also include part of the transmembrane region or part of the cytoplasmic domain or other sequences, provided that the soluble *ork* protein is capable of being secreted. Preferred soluble *ork* polypeptides include the signal sequence and entire extracellular domain (amino acids 1

to 745 of SEQ ID NO:1) or lack the signal sequence but contain the entire extracellular domain (amino acids 19 to 745 of SEQ ID NO:1).

Soluble *ork* polypeptides may be identified (and distinguished from their non-soluble membrane-bound counterparts) by separating intact cells which express the protein in question from the culture medium, e.g., by centrifugation, and assaying the medium (supernatant) for the presence of *ork*. The presence of *ork* in the medium indicates that the protein was secreted from the cells and thus is a soluble form. Soluble *ork* may be naturally-occurring forms of these proteins, such as those resulting from alternative splicing events. Alternatively, soluble fragments of *ork* proteins may be produced by recombinant DNA technology or otherwise isolated, as described below.

The use of soluble forms of *ork* is advantageous for certain applications. Purification of the proteins from recombinant host cells is facilitated, since the soluble proteins are secreted from the cells. The smaller soluble fragments may be advantageous for use in certain *in vitro* assays. The soluble *ork* polypeptides may be employed to competitively bind the ligand *in vivo*, thus inhibiting signal transduction activity via endogenous cell surface bound *ork* proteins. Further, soluble proteins are generally more suitable for intravenous administration and may exert their desired effect (e.g., binding a ligand) in the bloodstream.

Truncated *ork* proteins, including soluble polypeptides, may be prepared by any of a number of conventional techniques. In the case of recombinant proteins, a DNA fragment encoding a desired *ork* fragment may be subcloned into an expression vector. A desired DNA sequence may be chemially synthesized using known techniques. DNA fragments also may be produced by restriction endonuclease digestion of a full length cloned DNA sequence, and isolated by electrophoresis on agarose gels. Linkers containing restriction endonuclease cleavage site(s) may be employed to insert the desired DNA fragment into an expression vector, or the fragment may be digested at cleavage sites naturally present therein.

In another approach, enzymatic treatment (e.g., using Bal 31 exonuclease) may be employed to delete terminal nucleotides from a DNA fragment to obtain a fragment having a particular desired terminus. Among the commercially available linkers are those that can be ligated to the blunt ends produced by Bal 31 digestion, and which contain restriction endonuclease cleavage site(s). Alternatively, oligonucleotides that reconstruct the N- or C-terminus of a DNA fragment to a desired point may be synthesized. The oligonucleotide may contain a restriction endonuclease cleavage site upstream of the desired coding sequence and position an initiation codon (ATG) at the N-terminus of the coding sequence. The well known polymerase chain reaction procedure also may be employed to amplify a DNA sequence encoding a desired

protein fragment. 3' and 5' oligonucleotide primers that anneal to the *ork* DNA at the termini of a desired fragment are employed in the PCR reaction which is conducted using any suitable procedure, such as those described in Sarki et al., *Science* 239:487 (1988); in *Recombinant DNA Methodology*, Wu et al., eds., Academic Press Inc., San Diego (1989), pp. 189-196; and in *PCR Protocols: A Guide to Methods and Applications*, Innis et al., eds., Academic Press, Inc. (1990). An example of a suitable PCR procedure is as follows. All temperatures are in degrees centigrade. The following PCR reagents are added to a 0.5 ml Eppendorf microfuge tube: 10 µl of 10X PCR buffer (500 mM KCl, 100 mM Tris-HCl, pH 8.3 at 25°C, 25 mM $MgCl_2$, and 1 mg/ml gelatin) (Perkin-Elmer Cetus, Norwalk, CN), 8µl of a 2.5 mM solution containing each dNTP (2 mM dATP, 2 mM dCTP, 2 mM dGTP and 2 mM dTTP), 2.5 units (0.5 µl of standard 5000 units/ml solution) of *Taq* DNA polymerase (Perkins-Elmer Cetus), 1 ng of template DNA, 100 picomoles of each of the oligonucleotide primers, and water to a final volume of 100 µl. The final mixture is then overlaid with 100 µl parafin oil. PCR is carried out using a DNA thermal cycler (Ericomp, San Diego, CA). The template is denatured at 94° for 5 minutes and PCR is carried out for 25 cycles of amplification using a step program (denaturation at 94°, 1.5 minutes; annealing at 60°, 1 minute; extension at 72°, 1 minute).

The present invention also provides full length *ork* protein or antigenic fragments thereof that can act as immunogens to generate antibodies specific to the *ork* immunogens. Monoclonal antibodies specific for *ork* or antigenic fragments thereof are prepared by prodedures that include those described in Example 4. The above-described procedures for producing *ork* fragments may be employed in producing *ork* fragments for use as immunogens.

Expression of Recombinant *ork* Proteins

The present invention provides recombinant expression vectors to express DNA encoding the *ork* proteins of the present invention. The inventive recombinant expression vectors are replicable DNA constructs which contain a synthetic or cDNA-derived DNA sequence encoding an *ork* protein, operably linked to suitable transcriptional or translational regulatory elements. Examples of genetic elements having a regulatory role in gene expression include transcriptional promoters, operators or enhancers, a sequence encoding suitable mRNA ribosomal binding sites, and appropriate transcription and translation initiation and termination sequences. The ability to replicate in a host, usually conferred by an origin of replication, and a selection gene to facilitate recognition of transformants may additionally be incorporated. The regulatory elements employed in the expression vectors are generally

derived from mammalian, microbial, viral, or insect genes. Expression vectors derived from retroviruses also may be employed.

DNA regions are operably linked when they are functionally related to each other. A DNA sequence encoding *ork* is said to be operably linked to one or more of the above-described regulatory elements when the *ork* DNA sequence is transcribed, or the resulting mRNA is translated, under the control of the regulatory element(s).

Transformed host cells are cells which have been transformed or transfected with foreign DNA using recombinant DNA techniques. In the context of the present invention, the foreign DNA includes a sequence encoding the inventive *ork* protein. Host cells may be transformed for purposes of cloning or amplifying the foreign DNA, or may be transformed with an expression vector for production of the fusion protein under the control of appropriate promoters. Suitable host cells include prokaryotes, yeast, or higher eukaryotic cells. Appropriate cloning and expression vectors for use with bacterial, fungal, yeast, and mammalian cellular hosts are described by Pouwels et al. (*Cloning Vectors: A Laboratory Manual*, Elsevier, New York, 1985), the relevant disclosures of which is hereby incorporated by reference. Cell-free translation systems could also be employed to produce fusion protein using RNAs derived from the DNA constructs of the present invention.

Prokaryotes include gram negative or gram positive organisms. Prokaryotic expression vectors generally comprise one or more phenotypic selectable markers, for example a gene encoding proteins conferring antibiotic resistance or supplying an autotrophic requirement, and an origin of replication recognized by the host to ensure amplification within the host. Examples of suitable prokaryotic hosts for transformation include *E. coli*, bacilli such as *Bacillus subtilis*, *Salmonella typhimurium*, and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

Useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well-known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and pGEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. *E. coli* is typically transformed using derivatives of pBR322, a plasmid derived from an *E. coli* species (Bolivar et al., *Gene* 2:95, 1977). pBR322 contains genes for ampicillin and tetracycline resistance, providing simple means for identifying transformed cells.

Promoters commonly used in recombinant microbial expression vectors include the ß-lactamase (penicillinase) and lactose promoter system (Chang et al., *Nature*

275:615, 1978; and Goeddel et al., *Nature 281*:544, 1979), the tryptophan (trp) promoter system (Goeddel et al., *Nucl. Acids Res. 8*:4057, 1980; and EPA 36,776) and tac promoter (Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, p. 412, 1982). A particularly useful bacterial expression system employs the phage λ P_L promoter and cI857ts thermoinducible repressor. Plasmid vectors available from the American Type Culture Collection which incorporate derivatives of the λ P_L promoter include plasmid pHUB2, resident in *E. coli* strain JMB9 (ATCC 37092) and pPLc28, resident in E. coli RR1 (ATCC 53082).

The recombinant *ork* protein may also be expressed in yeast hosts, preferably from *Saccharomyces* species, such as *S. cerevisiae*. Yeast of other genera such as *Pichia* or *Kluyveromyces* may also be employed. Yeast vectors will generally contain an origin of replication from the 2μm yeast plasmid or an autonomously replicating sequence (ARS), a promoter, DNA encoding the *ork* protein, sequences for polyadenylation and transcription termination and a selection gene. Yeast vectors may include origins of replication and selectable markers permitting transformation of both yeast and *E. coli*, e.g., the ampicillin resistance gene of *E. coli* and the S. *cerevisiae* trp1 gene. The trp 1 gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, and a promoter derived from a highly expressed yeast gene to induce transcription of a structural sequence downstream. The presence of the trp1 lesion in the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan.

Suitable promoter sequences in yeast vectors include the promoters for metallothionein, 3-phosphoglycerate kinase (Hitzeman et al., *J. Biol. Chem. 255*:2073, 1980) or other glycolytic enzymes (Hess et al., *J. Adv. Enzyme Reg. 7*:149, 1968; and Holland et al., *Biochem. 17*:4900, 1978), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase and glucokinase. Examples of suitable vectors and promoters for use in yeast expression are further described in R. Hitzeman et al., EPA 73,657.

Preferred yeast vectors can be assembled using DNA sequences from pBR322 for selection and replication in *E. coli* (Amp^r gene and origin of replication) and yeast DNA sequences including a glucose-repressible ADH2 promoter and α–factor secretion leader. The ADH2 promoter has been described by Russell et al. (*J. Biol. Chem. 258*:2674, 1982) and Beier et al., (*Nature 300*:724, 1982). Advantageously, a DNA segment encoding a leader sequence functional in yeast is operably linked to the 5' end of the DNA encoding the *ork* protein. The encoded leader peptide promotes secretion of the *ork* protein from the host cell and is generally cleaved from the *ork*

protein upon secretion. As one example, the yeast α-factor leader, which directs secretion of heterologous proteins, can be inserted between the promoter and the structural gene to be expressed. See, e.g., Kurjan et al., *Cell 30*:922, 1982; and Bitter et al., *Proc. Natl. Acad. Sci. USA 81*:5330, 1984. The leader sequence may be modified to contain, near its 3' end, one or more useful restriction sites to facilitate fusion of the leader sequence to foreign genes.

Suitable yeast transformation protocols are known to those of skill in the art. An exemplary technique is described by Hinnen et al., *Proc. Natl. Acad. Sci. USA 75*:1929, (1978), selecting for Trp^+ transformants in a selective medium consisting of 0.67% yeast nitrogen base, 0.5% casamino acids, 2% glucose, 10μg/ml adenine and 20 μg/ml uracil. Host strains transformed by vectors comprising the above-described ADH2 promoter may be grown for expression in a rich medium consisting of 1% yeast extract, 2% peptone, and 1% glucose supplemented with 80 μg/ml adenine and 80 μg/ml uracil. Derepression of the ADH2 promoter occurs upon exhaustion of medium glucose. Crude yeast supernatants are harvested by filtration and held at 4°C prior to further purification.

Various mammalian or insect cell culture systems can be employed to express recombinant protein. Baculovirus systems for production of heterologous proteins in insect cells are reviewed by Luckow and Summers, *Bio/Technology 6*:47 (1988). Established cell lines of mammalian origin may be employed. Examples of suitable mammalian host cell lines include the COS-7 line of monkey kidney cells (ATCC CRL 1651), described by Gluzman (*Cell 23*:175, 1981), CV-1 cells (ATCC CCL 70) also derived from monkey kidney, L cells, C127, 3T3, Chinese hamster ovary (CHO), HeLa and BHK cell lines. Mammalian expression vectors may comprise non-transcribed elements such as an origin of replication, a suitable promoter and enhancer linked to the gene to be expressed, and other 5' or 3' flanking nontranscribed sequences, and 5' or 3' nontranslated sequences, such as necessary ribosome binding sites, a poly-adenylation site, splice donor and acceptor sites, and transcriptional termination sequences.

The transcriptional and translational control sequences in expression vectors to be used in transforming vertebrate cells may be provided by viral sources. For example, commonly used promoters and enhancers are derived from Polyoma, Adenovirus 2, Simian Virus 40 (SV40), and human cytomegalovirus. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early and late promoter, enhancer, splice, and polyadenylation sites may be used to provide the other genetic elements required for expression of a heterologous DNA sequence. The early and late promoters are particularly useful because both are obtained easily from the virus as a fragment which also contains the SV40 viral origin or replication (Fiers et al.,

Nature 273:113, 1978). Smaller or larger SV40 fragments may also be used, provided the approximately 250 bp sequence extending from the Hind III site toward the *BglI* site located in the viral origin of replication is included. Exemplary vectors can be constructed as disclosed by Okayama and Berg (*Mol. Cell. Biol. 3*:280, 1983). A useful system for stable high level expression of mammalian receptor cDNAs in C127 murine mammary epithelial cells can be constructed substantially as described by Cosman et al. (*Mol. Immunol. 23*:935, 1986). Other expression vectors for use in mammalian host cells are derived from retroviruses.

Producing and Purifying the *ork* Protein

The present invention provides substantially homogeneous *ork* protein, which may be produced by recombinant expression systems as described above or purified from naturally occurring cells. The ork protein is purified to substantial homogeneity, as indicated by a single protein band upon analysis by SDS-polyacrylamide gel electrophoresis (SDS-PAGE).

In one embodiment of the present invention, *ork* is purified from a cellular source using any suitable protein purification technique. The tissues identified in example 3 as containing *ork* RNA (preferably placental or lung tissue) from a mammalian species of interest may be employed as sources of *ork*, for example.

An alternative process for producing the recombinant *ork* protein of the present invention comprises culturing a host cell transformed with an expression vector comprising a DNA sequence that encodes said *ork* protein under conditions that promote expression of the *ork* protein, which is then purified from culture media or cell extracts. Any suitable purification process may be employed, with the procedure of choice varying according to such factors as the type of host cells and whether or not the desired protein is secreted from the host cells. The fusion protein will be secreted into the culture medium when it is initially fused to a signal sequence or leader peptide operative in the host cells, or when the protein comprises soluble forms of the *ork* polypeptides.

For example, supernatants from expression systems which secrete recombinant protein into the culture medium can be first concentrated using a commercially available protein concentration filter, e.g., an Amicon or Millipore Pellicon ultrafiltration unit. Following the concentration step, the concentrate can be applied to a suitable purification matrix. For example, an immunoaffinity column comprising antibodies directed against *ork* and bound to a suitable support may be employed. A monoclonal antibody specific for *ork* may be prepared as described in example 4. Alternatively, an anion exchange resin can be employed, for example, a matrix or substrate having pendant diethylaminoethyl (DEAE) groups. The matrices can be acrylamide, agarose,

dextran, cellulose or other types commonly employed in protein purification. Alternatively, a cation exchange step can be employed. Suitable cation exchangers include various insoluble matrices comprising sulfopropyl or carboxymethyl groups. Sulfopropyl groups are preferred. One or more reversed-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, e.g., silica gel having pendant methyl or other aliphatic groups, can be employed to further purify *ork*.

Recombinant protein produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more concentration, salting-out, aqueous ion exchange or size exclusion chromatography steps. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of recombinant fusion proteins can disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Fermentation of yeast which express *ork* as a secreted protein greatly simplifies purification. Secreted recombinant protein resulting from a large-scale fermentation can be purified by methods analogous to those disclosed by Urdal et al. (*J. Chromatog.* 296:171, 1984), involving two sequential, reversed-phase HPLC steps for purification of a recombinant protein on a preparative HPLC column.

Some or all of the foregoing purification steps, in various combinations, can be employed to provide an essentially homogeneous recombinant protein. Recombinant cell culture enables the production of the *ork* protein free of those contaminating proteins which may be normally associated with *ork* as it is found in nature, e.g., in cells, cell exudates or body fluids. The foregoing purification procedures are among those that may be employed to purify non-recombinant *ork* proteins of the present invention as well.

Variants and Derivatives of *ork*

Variants and derivatives of native *ork* proteins that retain the desired biological activity are also within the scope of the present invention. An *ork* variant, as referred to herein, is a polypeptide substantially homologous to a native *ork*, but which has an amino acid sequence different from that of native *ork* (human, murine or other mammalian species) because of one or a plurality of deletions, insertions or substitutions.

The variant amino acid sequence preferably is at least 80% identical to a native *ork* amino acid sequence, most preferably at least 90% identical. When the variant *ork* protein comprises extracellular, transmembrane and cytoplasmic domains, these percent identities apply to the entire sequence and also to the extracellular domain when taken

alone. Since the cytoplasmic domains are relatively conserved among members of the RTK family, the percent identity of the extracellular domain is important in identifying a variant as an *ork* protein.

To illustrate this point, a comparison of the full length *ork* and *tie* amino acid sequences (aligned in Figure 5) reveals 76% identity for the cytoplasmic domains, whereas the percent identity drops to 47.5% for the full length sequences as a whole. The percent identity drops to 33% when just the extracellular domains are compared. A DNA probe corresponding to the extracellular region of *tie* would not hybridize to *ork* DNA under moderately stringent hybridization conditions.

The degree of homology (percent identity) may be determined, for example, by comparing sequence information using the GAP computer program, version 6.0 described by Devereux et al. (*Nucl. Acids Res.* 12:387, 1984) and available from the University of Wisconsin Genetics Computer Group (UWGCG). The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443, 1970), as revised by Smith and Waterman (*Adv. Appl. Math* 2:482, 1981). Briefly, the GAP program defines similarity as the number of aligned symbols (i.e., nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. The preferred default parameters for the GAP program include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) for nucleotides, and the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res.* 14:6745, 1986, as described by Schwartz and Dayhoff, eds., *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, pp. 353-358, 1979; (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3) no penalty for end gaps.

Alterations of the native amino acid sequence may be accomplished by any of a number of known techniques. Mutations can be introduced at particular loci by synthesizing oligonucleotides containing a mutant sequence, flanked by restriction sites enabling ligation to fragments of the native sequence. Following ligation, the resulting reconstructed sequence encodes an analog having the desired amino acid insertion, substitution, or deletion.

Alternatively, oligonucleotide-directed site-specific mutagenesis procedures can be employed to provide an altered gene having particular codons altered according to the substitution, deletion, or insertion required. Exemplary methods of making the alterations set forth above are disclosed by Walder et al. (*Gene* 42:133, 1986); Bauer et al. (*Gene* 37:73, 1985); Craik (*BioTechniques*, January 1985, 12-19); Smith et al. (*Genetic Engineering: Principles and Methods*, Plenum Press, 1981); and U.S. Patent Nos. 4,518,584 and 4,737,462, which are incorporated by reference herein.

Variants may comprise conservatively substituted sequences, meaning that a given amino acid residue is replaced by a residue having similar physiochemical characteristics. Examples of conservative substitutions include substitution of one aliphatic residue for another, such as Ile, Val, Leu, or Ala for one another, or substitutions of one polar residue for another, such as between Lys and Arg; Glu and Asp; or Gln and Asn. Other such conservative substitutions, for example, substitutions of entire regions having similar hydrophobicity characteristics, are well known.

ork also may be modified to create *ork* derivatives by forming covalent or aggregative conjugates with other chemical moieties, such as glycosyl groups, lipids, phosphate, acetyl groups and the like. Covalent derivatives of *ork* may be prepared by linking the chemical moieties to functional groups on *ork* amino acid side chains or at the N-terminus or C-terminus of an *ork* polypeptide or the extracellular domain thereof. Other derivatives of *ork* within the scope of this invention include covalent or aggregative conjugates of *ork* or its fragments with other proteins or polypeptides, such as by synthesis in recombinant culture as N-terminal or C-terminal fusions. For example, the conjugate may comprise a heterologous signal or leader polypeptide sequence at the N-terminus of an *ork* polypeptide. Examples of such signal peptides are the α-factor leader of *Saccharomyces*; the signal sequence for interleukin-7 (IL-7) described in United States Patent 4,965,195; or the signal sequence for interleukin-2 receptor described in United States Patent Application 06/626,667 filed on July 2, 1984. The signal or leader peptide co-translationally or post-translationally directs transfer of the conjugate from its site of synthesis to a site outside of the cell membrane or cell wall.

ork polypeptide fusions can comprise peptides added to facilitate purification and identification of *ork* . Such peptides include, for example, poly-His or the antigenic identification peptides described in U.S. Patent No. 5,011,912 and in Hopp et al., *Bio/Technology* 6:1204, 1988. One such peptide is the FLAG® peptide, Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys (DYKDDDDK), which is highly antigenic and provides an epitope reversibly bound by a specific monoclonal antibody enabling rapid assay and facile purification of expressed recombinant protein. This sequence is also specifically cleaved by bovine mucosal enterokinase at the residue immediately following the Asp-Lys pairing. Fusion proteins capped with this peptide may also be resistant to intracellular degradation in *E. coli.* A murine hybridoma designated 4E11 produces a monoclonal antibody that binds the peptide DYKDDDDK in the presence of certain divalent metal cations (as described in U.S. Patent 5,011,912) and has been deposited with the American Type Culture Collection under accession no HB 9259.

The present invention further includes *ork* polypeptides with or without associated native-pattern glycosylation. *ork* expressed in yeast or mammalian

expression systems (e.g., COS-7 cells) may be similar to or significantly different from a native *ork* polypeptide in molecular weight and glycosylation pattern, depending upon the choice of expression system. Expression of *ork* polypeptides in bacterial expression systems, such as *E. coli*, provides non-glycosylated molecules.

DNA constructs that encode various additions or substitutions of amino acid residues or sequences, or deletions of terminal or internal residues or sequences not needed for biological activity or binding, can be prepared. For example, N-glycosylation sites in the *ork* extracellular domain can be modified to preclude glycosylation while allowing expression of a homogeneous, reduced carbohydrate analog using yeast expression systems. N-glycosylation sites in eukaryotic polypeptides are characterized by an amino acid triplet Asn-X-Y, wherein X is any amino acid except Pro and Y is Ser or Thr. Appropriate modifications to the nucleotide sequence encoding this triplet will result in substitutions, additions or deletions that prevent attachment of carbohydrate residues at the Asn side chain. Known procedures for inactivating N-glycosylation sites in proteins include those described in U.S. Patent 5,071,972 and EP 276,846. In another example, sequences encoding Cys residues that are not essential for biological activity can be altered to cause the Cys residues to be deleted or replaced with other amino acids, preventing formation of incorrect intramolecular disulfide bridges upon renaturation. Other variants are prepared by modification of adjacent dibasic amino acid residues to enhance expression in yeast systems in which KEX2 protease activity is present. EP 212,914 discloses the use of site-specific mutagenesis to inactivate KEX2 protease processing sites in a protein.

Naturally occurring *ork* variants are also encompassed by the present invention. Examples of such variants are proteins that result from alternative mRNA splicing events (since *ork* is encoded by a multi-exon gene) or from proteolytic cleavage of the *ork* protein, provided the desired biological activity (e.g., binding to an anti-*ork* antibody or to the ligand) is retained. Alternative splicing of mRNA may yield a truncated but biologically active *ork* protein, such as a naturally occurring soluble form of the protein, for example. An alternative splicing event in *tie*, an RTK of the same subgroup as *ork*, yielded a protein lacking the first of three fibronectin type III repeats. Variations attributable to proteolysis include, for example, differences in the N- or C-termini upon expression in different types of host cells, due to proteolytic removal of one or more terminal amino acids (which may occur intracellularly or during purification). Varying N-termini may also result from cleavage of the signal peptide in certain host cells at a point other than between amino acids 18 and 19 of the disclosed sequence.

In certain host cells, post-translational processing will remove the methionine residue encoded by an initiation codon, whereas the methionine residue will remain at

the N-terminus of proteins produced in other host cells. The N-terminal amino acid may, for example, be any of the amino acids at positions 1 to 5 of SEQ ID NO:1 (for proteins comprising a signal peptide) or 19-23 (for the mature protein). The C-terminus may be truncated deliberately during expression vector construction (e.g., in constructing vectors encoding soluble proteins as described above) or as a result of differential processing which may remove up to about five C-terminal amino acids, for example.

Due to the known degeneracy of the genetic code wherein more than one codon can encode the same amino acid, a DNA sequence may vary from that presented in SEQ ID NO:1, and still encode an *ork* protein having the amino acid sequence of SEQ ID NO:1. Such variant DNA sequences may result from silent mutations (e.g., occurring during PCR amplification), and may be the product of deliberate mutagenesis of a native sequence.

Nucleic acid sequences within the scope of the present invention include isolated DNA and RNA sequences that hybridize to the *ork* nucleotide sequences disclosed herein under conditions of moderate or severe stringency, and which encode biologically active *ork*. Moderate stringency hybridization conditions refer to conditions described in, for example, Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 2 ed. Vol. 1, pp. 1.101-104, Cold Spring Harbor Laboratory Press, (1989). Conditions of moderate stringency, as defined by Sambrook et al., include use of a prewashing solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0) and hybridization conditions of about 55°C, 5 X SSC, overnight. Conditions of severe stringency include higher temperatures of hybridization and washing. The skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as the length of the probe.

The present invention thus provides isolated DNA sequences encoding biologically active *ork* , selected from: (a) DNA derived from the coding region of a native mammalian *ork* gene (e.g., cDNA derived from the coding region of the human *ork* cDNA sequence presented in SEQ ID NO:1); (b) DNA capable of hybridizing under moderately stringent conditions to a DNA derived from the extracellular region of the sequence presented in SEQ ID NO:1, and which encodes biologically active *ork* ; and (c) DNA which is degenerate as a result of the genetic code to a DNA defined in (a) or (b) and which encodes biologically active *ork*.

ork polypeptides in the form of oligomers such as dimers or trimers are within the scope of the present invention. Oligomers may be linked by disulfide bonds formed between cysteine residues on different *ork* polypeptides. In one embodiment of the invention, an *ork* dimer is created by fusing *ork* to the Fc region of an antibody (IgG1). The Fc polypeptide preferably is fused to the C-terminus of a soluble *ork* (comprising

only the extracellular domain). A gene fusion encoding the *ork* fusion protein is inserted into an appropriate expression vector. The *ork* fusion proteins are allowed to assenble much like antibody molecules, whereupon interchain disulfide bonds form between Fc polypeptides, yielding divalent *ork*. If fusion proteins are made with both heavy and light chains of an antibody, it is possible to form an *ork* oligomer with as many as four *ork* extracellular regions. Alternatively, one can link two soluble ork domains with a peptide linker such as the $Gly_4SerGly_5Ser$ linker sequence described in United States Patent 5,073,627. A fusion protein comprising two or more *ork* polypeptides (with or without peptide spacers) may be produced by recombinant DNA technology.

The present invention further provides antisense or sense oligonucleotides comprising a single-stranded nucleic acid sequence (either RNA or DNA) capable of binding to target *ork* mRNA (sense) or *ork* DNA (antisense) sequences. Antisense or sense oligonucleotides, according to the present invention, comprise a fragment of the coding region of *ork* cDNA. Such a fragment generally comprises at least about 14 nucleotides, preferably from about 14 to about 30 nucleotides. The ability to create an antisense or a sense oligonucleotide, based upon a cDNA sequence for a given protein is described in, for example, Stein and Cohen, *Cancer Res.* 48:2659, 1988 and van der Krol et al., *BioTechniques* 6:958, 1988.

Binding of antisense or sense oligonucleotides to target nucleic acid sequences results in the formation of duplexes that block translation (RNA) or transcription (DNA) by one of several means, including enhanced degradation of the duplexes, premature termination of transcription or translation, or by other means. The antisense oligonucleotides thus may be used to block expression of *ork* proteins. Antisense or sense oligonucleotides further comprise oligonucleotides having modified sugar-phosphodiester backbones (or other sugar linkages, such as those described in WO91/06629) and wherein such sugar linkages are resistant to endogenous nucleases. Such oligonucleotides with resistant sugar linkages are stable *in vivo* (i.e., capable of resisting enzymatic degradation) but retain sequence specificity to be able to bind to target nucleotide sequences. Other examples of sense or antisense oligonucleotides include those oligonucleotides which are covalently linked to organic moieties, such as those described in WO 90/10448, and other moieties that increase affinity of the oligonucleotide for a target nucleic acid sequence, such as poly-(L-lysine). Further still, intercalating agents, such as ellipticine, and alkylating agents or metal complexes may be attached to sense or antisense oligonucleotides to modify binding specificities of the antisense or sense oliginucleotide for the target nucleotide sequence.

Antisense or sense oligonucleotides may be introduced into a cell containing the target nucleic acid sequence by any gene transfer method, including, for example,

$CaPO_4$-mediated DNA transfection, electroporation, or other gene transfer vectors such as Epstein-Barr virus. Antisense or sense oligonucleotides are preferably introduced into a cell containing the target nucleic acid sequence by insertion of the antisense or sense oligonucleotide into a suitable retroviral vector, then contacting the cell with the retrovirus vector containing the inserted sequence, either *in vivo* or *ex vivo*. Suitable retroviral vectors include, but are not limited to, the murine retrovirus M-MuLV, N2 (a retrovirus derived from M-MuLV), or or the double copy vectors designated DCT5A, DCT5B and DCT5C (see PCT Application US 90/02656). Alternatively, other promotor sequences may be used to express the oligonucleotide.

Sense or antisense oligonucleotides may also be introduced into a cell containing the target nucleotide sequence by formation of a conjugate with a ligand binding molecule, as described in WO 91/04753. Suitable ligand binding molecules include, but are not limited to, cell surface receptors, growth factors, other cytokines, or other ligands that bind to cell surface receptors.

Alternatively, a sense or an antisense oligonucleotide may be introduced into a cell containing the target nucleic acid sequence by formation of an oligonucleotide-lipid complex, as described in WO 90/10448. The sense or antisense oligonucleotide-lipid complex is preferably dissociated within the cell by an endogenous lipase.

Uses of Ork Proteins

One use of *ork* is as a research tool for identifying the ligand that binds thereto and studying the biological effects of ligand binding. The *ork* polypeptides of the present invention also may be employed in *in vitro* assays for detection of *ork* or its ligand or the interactions thereof.

The *ork* polypeptides of the present invention can be used in a binding assay to detect cells expressing a ligand for *ork*. For example, *ork* or an extracellular domain or a fragment thereof can be conjugated to a detectable moiety such as ^{125}I. Radiolabeling with ^{125}I can be performed by any of several standard methodologies that yield a functional ^{125}I *ork* molecule labeled to high specific activity. Alternatively, another detectable moiety such as an enzyme that can catalyze a colorometric or fluorometric reaction, biotin or avidin may be used. Cells to be tested for *ork* ligand expression can be contacted with the labeled *ork*. After incubation, unbound labeled *ork* is removed and binding is measured using the detectable moiety.

Soluble *ork* polypeptides may be employed to competitively bind the ligand *in vivo*, thus inhibiting signal transduction activity via endogenous cell surface bound *ork* proteins. Further, soluble proteins are generally more suitable for intravenous administration and may exert their desired effect (e.g. binding a ligand) in the bloodstream. Soluble *ork* proteins comprising only the extracellular (ligand-binding)

domain lack the tyrosine kinase domain which is located within the cytoplasmic domain.

The *ork* polypeptides disclosed herein are also useful in generating antibodies specific to the *ork* immunogens. Monoclonal antibodies specific for *ork* or antigenic fragments thereof may be prepared by procedures that include those described in Example 4. The monoclonal antibodies can be attached to insoluble support materials for use in immunoaffinity column purification of *ork* proteins. Other uses for the antibodies include detecting *ork* proteins in *in vitro* assays and identifying and purifying additional *ork* polypeptides such as variants (e.g., from alternative splicing events) that comprise the region from which the immunogen was derived.

EXAMPLE 1: PCR-Based Cloning of Human *ork* cDNA

Degenerate oligonucleotide primers were synthesized corresponding to the conserved sequences HRDLAA (TK-1; sense orientation) and SDVWS (TK-2; antisense orientation) contained within the kinase domain of all RTKs (Hanks, *et al.*, 1988). The oligonucleotides corresponding to the conserved motifs HRDLAA and SDVWS were
(5'-CAC$^{C}/_{A}$AG$^{G}/_{A}$GA$^{C}/_{T}$$^{C}/_{T}$TGGC$^{A}/_{T}$GC-3') and
(5'-A$^{G}/_{A}$GACCA$^{A/T}/_{C/G}$AC$^{G}/_{A}$TC$^{G}/_{A}$CT-3'), respectively. Both oligonucleotides were 32-fold degenerate and contained recognition sequences for *Xho* I. The HRDLAA sequence was chosen to favor RTKs over *src*-family tyrosine kinases that have the sequence HRDLRA (Hanks *et al.*, 1988).

Single-stranded cDNA was synthesized from human placental polyadenylated RNA by standard methods. The single stranded cDNA was used as template for a PCR reaction using the conditions of Wilks (1989). The two degenerate oligonucleotides were used as primers in the PCR reaction. Amplified PCR reaction products of about 200 base pairs were digested with *Xho* I and ligated into an *Xho* I-digested plasmid vector desinated pBLUESCRIPT® SK. This vector, available from Stratagene Cloning Systems, La Jolla, California, is replicable in *E. coli* and contains a polylinker segment that includes 21 unique restriction sites, one of which is *Xho* I. The ligation mixture was transformed into *E. coli* cells by conventional procedures.

Recombinant plasmids were recovered from the transformed *E. coli* cells and the nucleotide sequence of the DNA insert in a number of individual clones was determined. Among the known RTKs that were represented in the PCR-library were c-*fms*, JTK-4 (Partanen *et al. Proc. Natl. Acad. Sci. USA*, 87:8913-17, 1990), *kdr* (Terman *et al., Oncogene*, 6:1677-1683, 1991), *eph* (Hirai *et al.*, 1987 *supra*), and *flt*-4 (Aprelikova *et al. Cancer Res.*, 52:746-748, 1992). However, one clone, designated HPK-6, contained a novel sequence. The approximately 200bp cDNA

insert of HPK-6 was isolated and radiolabeled with ^{32}P by conventional techniques for use as a probe to isolate longer cDNA sequences encoding the novel protein.

An oligo-dT-primed human placental cDNA library in plasmid pDC302 has been previously described (Larsen *et al., J. Exp. Med.*, 172:1559-1570, 1990). Briefly, total cell RNA was isolated from whole fresh placental tissue and polyadenylated RNA was prepared by chromatography on oligo(dT)-cellulose. Double-stranded, oligo(dT)-primed cDNA was prepared with a commercial kit (Amersham Corp., Arlington Heights, IL). The resulting cDNA was size fractionated by chromatography on Sephacryl S-1000 (Pharmacia Fine Chemicals, Piscataway, NJ) in 0.5 M sodium acetate. The excluded cDNA was cloned into the *BglII* site of the mammalian expresion vector, pDC302 (described below) by an adaptor method similar to that described by Haymerle et al. (*Nucl. Acids Res.* 14:8615, 1986). Briefly, noncomplementary oligonucleotides of the sequence 5'-GATCTTGGAACGAGACGACCTGCT and 5'-AGCAGGTCGTCTCGTTCCAA synthesized on a DNA synthesizer (model 380A; Applied Biosystems, Foster City, CA) were annealed and ligated in separate reactions to either cDNA or *BglII* cut vector. Nonligated oligonucleotides were separated from cDNA or vector by chromatography over Sepharose CL-2B (Pharmacia Fine Chemicals) at 65°C in 10mM Tris (pH 8.0), 0.1mM EDTA. 5 ng of adaptored vector was ligated to adaptored cDNA in 10-µl reactions containing 50 mM sodium chloride, 50 mM Tris-HCl (pH 7.5), 10 mM magnesium chloride, 1 mM spermidine, 0.5 mM ATP, 0.1 U/µl T4 polynucleotide kinase and 0.4 U/µl T4 DNA ligase for 30 min at 37°C. Reactions were then desalted by drop dialysis on VSWP 013 filters (Millipore Corp., Bedford, MA) against distilled water for 40 min immediately before electroporation into *Escherichia coli* strain DH5α. Transformants were obtained with an average cDNA insert size of 1.6 kb.

The pDC302 expression vector employed in preparing this cDNA library has been described by Mosley et al. (*Cell* 59:335, 1989). pDC302 is an expression vector for use in mammalian host cells, but also replicates in *E. coli*.

pDC302 was assembled from pDC201 (Sims et al., *Science* 241:585, 1988), SV40 and cytomegalovirus DNA and comprises, in order with the direction of transcription from the origin of replication: (1) SV40 sequences from coordinates 5171-270 including the origin of replication, enhancer sequences and early and late promoters; (2) human cytomegalovirus sequences including the promoter and enhancer regions (nucleotides 671 to 63 from the sequence published by Boechart et al. (*Cell* 41:521, 1985); (3) adenovirus-2 sequences containing the first exon and part of the intron between the first and second exons of the tripartite leader, the second exon and part of the third exon of the tripartite leader; and (4) a multiple cloning site (MCS) containing sites for XhoI, Asp718, SmaI, NotI and BglII; (5) SV40 sequences from

coordinates 4127-4100 and 2770-2533 that include the polyadenylation and termination signals for early transcription; (6) sequences derived from pBR322 and virus-associated sequences VAI and VAII of pDC201, with adenovirus sequences 10532-11156 containing the VAI and VAII genes, followed by pBR322 sequences from 4363-2486 and 1094-375 containing the ampicillin resistance gene and origin of replication.

Miniprep DNA isolated from individual pools of colonies from the above-described placental cDNA library was digested with *Bgl* II (which excises the cDNA inserts) and screened for hybridization to the HPK-6-derived probe. A clone comprising a 2.3 Kb cDNA insert was isolated from one pool (#398) and the insert was radiolabeled for use in subsequent library screens.

A random-primed cDNA library in λgt10 was generated from poly(A)$^+$ human placental RNA and screened for hybridization to the 2.3 kb cDNA probe. One clone (32-1), about 4.0 Kb in length, contained an entire coding region, as well as 5' and 3' untranslated sequences.

A DNA sequence for human *ork* is presented in SEQ ID NO:1, along with the amino acid sequence encoded thereby. The SEQ ID NO:1 sequence was derived by combining DNA sequencing information from the fully-sequenced 2.3kb clone with that obtained for the 4.0 kb clone (for which certain portions overlapping with the 2.3 kb clone were not fully sequenced). Beginning with an ATG codon at nucleotides 149-151 there is an open reading frame extending for 1124 amino acids. Several lines of evidence suggest that this methionine codon is the initiating codon. The ATG codon is in a proper context (Kozak, 1984), and is followed by a hydrophobic stretch of amino acids that resembles a leader sequence (amino acids 1-18 of SEQ ID NO:1, i.e., methionine through glycine). There are two in-frame termination codons upstream of the ATG codon. The protein comprises an N-terminal extracellular (ligand-binding) domain (amino acids 19-745), followed by a transmembrane region comprising amino acids 746-772, and a C-terminal cytoplasmic domain (which contains the tyrosine kinase domain responsible for the tyrosine phosphorylating activity of the protein) comprising amino acids 773-1124.

The 2.3 kb clone encodes a C-terminal fragment of *ork* , including a small C-terminal portion of the extracellular domain, followed by the complete transmembrane and cytoplasmic domains of the protein (amino acids 698-1124). The 2.3 kb *ork* cDNA extends from nucleotide 2240 to the 3' end of the SEQ ID NO:1 DNA sequence and contains additional 3' non-coding sequences (i.e., the 2.3 kb clone has a longer 3' non-coding region than does the 4.0 kb clone depicted in SEQ ID NO:1). Since the 2.3 kb clone lacks most of the extracellular domain, the tyrosine kinase encoded thereby is not expected to bind a ligand.

The 4.0 kb *ork* cDNA in plasmid pBLUESCRIPT®SK in *E. coli* strain DH5α was deposited with the American Type Culture collection on May 28, 1992 and was given accession number ATCC 69003. The strain deposit was made under the terms of the Budapest Treaty.

EXAMPLE 2: Characterization of *ork* Gene Product

Rabbit polyclonal antiserum was generated against a peptide consisting of the carboxyl-terminal 21 amino acids of the *ork* protein cytoplasmic domain (SEQ ID NO:1 sequence) conjugated to ovalbumin (Kitagawa and Aikawa, *J. Biochem.*79:233-236, 1976). Animals were immunized with 100 µg of the conjugate in complete Freund's adjuvant, followed by three boosts in incomplete adjuvant, after which time test bleeds were taken and assayed for immunoreactivity towards the peptide immobilized on nitrocellulose. Bleeds that showed anti-peptide activity were used in the following immunoprecipitation experiment.

To examine the protein encoded by the *ork* cDNA, the full length cDNA was excised from λgt10 by *Not* I digestion and ligated into a *Not* I-digested expression plasmid pDC302 (described above). *Not* I cleaves pDC302 at a unique site within a multiple cloning site. The resulting recombinant vectors were transfected into COS-7 cells and 3 days later the cells were labelled with ^{35}S-met/cys and detergent solubilized. The ^{35}S-labeled lysates were incubated with 5 µl of the anti-peptide serum for 1 hour at 4°C, after which time 100 µl of pre-swollen proteinA-sepharose was added and the lysates were incubated for an additional hour at 4°C. The immune complexes were washed 3 times with PBS/1% Triton X-100, boiled in SDS-PAGE sample buffer, and analyzed by SDS/polyacrylamide gel electrophoresis. The gel was treated with Amplify (Amersham, Arlington Heights, IL) and exposed to X-Ray film.

As shown in Figure 3 (panel A), the immune serum, but not the pre-immune serum, precipitated an approximately 140 Kd protein in the lysates from the *ork*-transfected cells, but not in lysates from control cells. Preincubation of the immune serum with the peptide specifically blocked the precipitation of the 140 Kd protein.

A hallmark of protein tyrosine kinases is their ability to auto-phosphorylate *in vitro* in immune complexes (Ullrich and Schlessinger, 1990)). We tested the ability of the *ork* protein to autophosphorylate using the rabbit serum described above. *Ork*- or mock-transfected COS-7 cells were washed 1X in cold PBS/0.1mM sodium orthovanadate and lysed at 10^7 cells/ml in a buffer containing: 25mM TRIS pH 8/150mM NaCl/1mM EGTA/1mM DTT/1% NP-40/0.1mM sodium orthovanadate/1mM PMSF/10 µg/ml leupeptin/10 µg/ml pepstatin A. Lysates were incubated with 5 µl of the above-described rabbit anti-peptide serum or normal rabbit

serum for 2 hours at 4°C with 15 µl (packed volume) protein A-sepharose. Immune complexes were washed 3X with 25mM TRIS pH 8/150mM NaCl/1% Triton X-100/0.1% SDS/1% sodium deoxycholate/1mM DTT/0.1mM sodium orthovanadate/1mM PMSF/ 10 µg/ml leupeptin, followed by three washes in 20mM HEPES pH 7.4/10mM $MnCl_2$/5mM $MgCl_2$. *In vitro* phosphorylation reactions were initiated by suspending the beads to a final volume of 50 µl in a mixture containing [γ-^{32}P]ATP (25mM)/10mM $MnCl_2$/5mM $MgCl_2$/20mM HEPES pH 7.4. After incubation for 30 minutes at 37°C the beads were washed with 20mM HEPES pH 7.4/10mM ATP/5mM EGTA, boiled in SDS-PAGE sample buffer and separated on 8-16% SDS/polyacrylamide gels followed by autoradiography. For phosphoaminoacid analysis, separated proteins were electrophoretically transferred to membranes and regions containing ^{32}P-labelled protein were excised and incubated with 6N HCl for 1 hour at 110°C. Hydrolysates were analysed by two dimensional thin-layer electrophoresis (Cooper *et al.*, 1983) and autoradiography.

As shown in Figure 3, panel B, immunoprecipitated *ork* was capable of auto-phosphorylation. Phospho-amino acid analysis demonstrated that only phospho-tyrosine was present in the immunoprecipitates (Fig. 3, panel C). We also examined whether the *ork* expressed in the COS cells was phosphorylated on tyrosine residues. Lysates from *ork*- and control-transfected cells were immunoprecipated with the P1 (polyclonal) serum, separated by SDS-PAGE, transferred to filters and incubated with an anti-phosphotyrosine anti-serum. As shown in Figure 3, panel D, a 140 Kd protein containing phosphotyrosine was seen only in the lysates from *ork*-transfected cells that had been immunoprecipitated with the anti-*ork* serum. It is unclear from these data whether *ork* has a high level of intrinsic tyrosine kinase activity, whether the COS cells express a potential ligand for *ork* which is interacting with the *ork* protein on the COS cell surface, or whether *ork* is being phosphorylated by a tyrosine kinase expressed by COS cells.

EXAMPLE 3: Screening Human Tissues for *ork* Expression

RNAs from several human tissues were screened for *ork* transcripts. For *ork* expression, a filter containing 2µg of poly(A)$^+$ RNA from various human tissues (Clontech, San Diego, CA) was hybridized with an antisense RNA probe containing nucleotides 2240-3005 (outside of the kinase domain). As shown in Figure 4, a 4.5 kb mRNA was found at highest levels in placenta and lung, with lower levels detected in kidney and heart. Each of these tissues is highly vascularized, suggesting that *ork* may be expressed in endothelial cells. Consistent with this, umbilical cord vein endothelial cells (HUVECs) also expressed *ork* message. In addition to the tissues

listed in Figure 4, RNAs isolated from hematopoietic cell lines and tissues were surveyed, but *ork* mRNA was not detected.

EXAMPLE 4: Monoclonal Antibodies Directed Against *ork*

This example illustrates the preparation of monoclonal antibodies to *ork* or immunogenic fragments thereof. Purified *ork* can be used to generate anti-*ork* monoclonal antibodies using conventional techniques, for example, those techniques described in U.S. Patent 4,411,993.

In one embodiment of the invention, the peptide used to generate rabbit polyclonal antiserum in example 2 is used to produce a monoclonal antibody. The peptide comprises the carboxy-terminal 21 amino acids of the *ork* protein (i.e., the last 21 amino acids of the cytoplasmic domain in the SEQ ID NO:1 sequence) conjugated to ovalbumin. Briefly, mice are immunized with the immunogenic peptide emulsified in complete Freund's adjuvant, and injected in amounts ranging from 10-100 μg subcutaneously or intraperitoneally. Ten to twelve days later, the immunized animals are boosted with additional immunogenic peptide emulsified in incomplete Freund's adjuvant. Mice are periodically boosted thereafter on a weekly to bi-weekly immunization schedule. Serum samples are periodically taken by retro-orbital bleeding or tail-tip excision for testing by dot blot assay or ELISA (Enzyme-Linked Immunosorbent Assay), for anti-*ork* antibodies.

Following detection of an appropriate antibody titer, positive animals are provided one last intravenous injection of the immunogen in saline. Three to four days later, the animals are sacrificed, spleen cells harvested, and spleen cells are fused to a murine myeloma cell line (e.g., NS1 or Ag 8.653). Fusions generate hybridoma cells, which are plated in multiple microtiter plates in a HAT (hypoxanthine, aminopterin and thymidine) selective medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

The hybridoma cells are screened by ELISA for reactivity against purified *ork* peptide by adaptations of the techniques disclosed in Engvall et al., *Immunochem.* 8:871, 1971 and in U.S. Patent 4,703,004. Positive hybridoma cells can be injected intraperitoneally into syngeneic BALB/c mice to produce ascites containing high concentrations of anti-*ork* monoclonal antibodies. Alternatively, hybridoma cells can be grown *in vitro* in flasks or roller bottles by various techniques. Monoclonal antibodies produced in mouse ascites can be purified by ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively, affinity chromatography based upon binding of antibody to protein A or protein G can also be used, as can affinity chromatography based upon binding to *ork*.

EXAMPLE 5: Cloning of Murine *ork* cDNA

Murine *ork* cDNA was isolated by cross-species hybridization using human *ork* cDNA as a probe. A murine cDNA library was prepared from 12 1/2 day old mouse embryos. The library was screened by conventional techniques using the radiolabeled 2.3 Kb human *ork* cDNA described in example 1 as a probe.

The DNA sequence of the cDNA insert of a hybridizing clone was determined, and the amino acid sequence encoded thereby is presented in SEQ ID NO:3 and in Figure 6. The murine *ork* sequence (designated "m"; in the top rows) is aligned with the corresponding portion of human *ork* ("h") in Figure 6. Identical amino acids are indicated with a line; conservative or similar amino acid changes are indicated with two dots or one dot, respectively. The percent similarity of these murine and human *ork* amino acid sequences is 99.2% and the percent identity is 97.5%, as determined by the GAP computer program.

The murine *ork* sequence corresponds to amino acids 886-1124 of human *ork*, a region falling within the cytoplasmic domain and extending to the C-terminus of the human *ork* protein. The murine cDNA may be used as a probe to screen the same or a different murine cDNA library to identify a full length murine *ork* clone. Alternatively, given the similarity of the human and murine sequences, murine libraries may be screened using a probe derived from the extracellular region of human *ork*.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Ziegler, Steven F.

(ii) TITLE OF INVENTION: NOVEL TYROSINE KINASE

(iii) NUMBER OF SEQUENCES: 3

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Immunex Corporation
(B) STREET: 51 University Street
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: US
(F) ZIP: 98101

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/905,600
(B) FILING DATE: 26-JUN-1992
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Seese, Kathryn A.
(B) REGISTRATION NUMBER: 32,172
(C) REFERENCE/DOCKET NUMBER: 2609

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (206) 587-0430
(B) TELEFAX: (206) 233-0644
(C) TELEX: 756822

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4138 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 149..3523

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTTCTGTGCT GTTCCTTCTT GCCTCTAACT TGTAAACAAG ACGTACTAGG ACGATGCTAA 60

TGGAAAGTCA CAAACCGCTG GGTTTTTGAA AGGATCCTTG GGACCTCATG CACATTTGTG 120

```
GAAACTGGAT GGAGAGATTT GGGGAAGC ATG GAC TCT TTA GCC AGC TTA GTT          172
                                Met Asp Ser Leu Ala Ser Leu Val
                                  1               5

CTC TGT GGA GTC AGC TTG CTC CTT TCT GGA ACT GTG GAA GGT GCC ATG          220
Leu Cys Gly Val Ser Leu Leu Leu Ser Gly Thr Val Glu Gly Ala Met
     10                  15                  20

GAC TTG ATC TTG ATC AAT TCC CTA CCT CTT GTA TCT GAT GCT GAA ACA          268
Asp Leu Ile Leu Ile Asn Ser Leu Pro Leu Val Ser Asp Ala Glu Thr
 25                  30                  35                  40

TCT CTC ACC TGC ATT GCC TCT GGG TGG CGC CCC CAT GAG CCC ATC ACC          316
Ser Leu Thr Cys Ile Ala Ser Gly Trp Arg Pro His Glu Pro Ile Thr
                 45                  50                  55

ATA GGA AGG GAC TTT GAA GCC TTA ATG AAC CAG CAC CAG GAT CCG CTG          364
Ile Gly Arg Asp Phe Glu Ala Leu Met Asn Gln His Gln Asp Pro Leu
             60                  65                  70

GAA GTT ACT CAA GAT GTG ACC AGA GAA TGG GCT AAA AAA GTT GTT TGG          412
Glu Val Thr Gln Asp Val Thr Arg Glu Trp Ala Lys Lys Val Val Trp
         75                  80                  85

AAG AGA GAA AAG GCT AGT AAG ATC AAT GGT GCT TAT TTC TGT GAA GGG          460
Lys Arg Glu Lys Ala Ser Lys Ile Asn Gly Ala Tyr Phe Cys Glu Gly
     90                  95                  100

CGA GTT CGA GGA GAG GCA ATC AGG ATA CGA ACC ATG AAG ATG CGT CAA          508
Arg Val Arg Gly Glu Ala Ile Arg Ile Arg Thr Met Lys Met Arg Gln
105                 110                 115                 120

CAA GCT TCC TTC CTA CCA GCT ACT TTA ACT ATG ACT GTG GAC AAG GGA          556
Gln Ala Ser Phe Leu Pro Ala Thr Leu Thr Met Thr Val Asp Lys Gly
                125                 130                 135

GAT AAC GTG AAC ATA TCT TTC AAA AAG GTA TTG ATT AAA GAA GAA GAT          604
Asp Asn Val Asn Ile Ser Phe Lys Lys Val Leu Ile Lys Glu Glu Asp
            140                 145                 150

GCA GTG ATT TAC AAA AAT GGT TCC TTC ATC CAT TCA GTG CCC CGG CAT          652
Ala Val Ile Tyr Lys Asn Gly Ser Phe Ile His Ser Val Pro Arg His
        155                 160                 165

GAA GTA CCT GAT ATT CTA GAA GTA CAC CTG CCT CAT GCT CAG CCC CAG          700
Glu Val Pro Asp Ile Leu Glu Val His Leu Pro His Ala Gln Pro Gln
    170                 175                 180

GAT GCT GGA GTG TAC TCG GCC AGG TAT ATA GGA GGA AAC CTC TTC ACC          748
Asp Ala Gly Val Tyr Ser Ala Arg Tyr Ile Gly Gly Asn Leu Phe Thr
185                 190                 195                 200

TCG GCC TTC ACC AGG CTG ATA GTC CGG AGA TGT GAA GCC CAG AAG TGG          796
Ser Ala Phe Thr Arg Leu Ile Val Arg Arg Cys Glu Ala Gln Lys Trp
                205                 210                 215

GGA CCT GAA TGC AAC CAT CTC TGT ACT GCT TGT ATG AAC AAT GGT GTC          844
Gly Pro Glu Cys Asn His Leu Cys Thr Ala Cys Met Asn Asn Gly Val
            220                 225                 230
```

```
TGC CAT GAA GAT ACT GGA GAA TGC ATT TGC CCT CCT GGG TTT ATG GGA      892
Cys His Glu Asp Thr Gly Glu Cys Ile Cys Pro Pro Gly Phe Met Gly
        235                 240                 245

AGG ACG TGT GAG AAG GCT TGT GAA CTG CAC ACG TTT GGC AGA ACT TGT      940
Arg Thr Cys Glu Lys Ala Cys Glu Leu His Thr Phe Gly Arg Thr Cys
    250                 255                 260

AAA GAA AGG TGC AGT GGA CAA GAG GGA TGC AAG TCT TAT GTG TTC TGT      988
Lys Glu Arg Cys Ser Gly Gln Glu Gly Cys Lys Ser Tyr Val Phe Cys
265                 270                 275                 280

CTC CCT GAC CCC TAT GGG TGT TCC TGT GCC ACA GGC TGG AAG GGT CTG     1036
Leu Pro Asp Pro Tyr Gly Cys Ser Cys Ala Thr Gly Trp Lys Gly Leu
                285                 290                 295

CAG TGC AAT GAA GCA TGC CAC CCT GGT TTT TAC GGG CCA GAT TGT AAG     1084
Gln Cys Asn Glu Ala Cys His Pro Gly Phe Tyr Gly Pro Asp Cys Lys
            300                 305                 310

CTT AGG TGC AGC TGC AAC AAT GGG GAG ATG TGT GAT CGC TTC CAA GGA     1132
Leu Arg Cys Ser Cys Asn Asn Gly Glu Met Cys Asp Arg Phe Gln Gly
        315                 320                 325

TGT CTC TGC TCT CCA GGA TGG CAG GGG CTC CAG TGT GAG AGA GAA GGC     1180
Cys Leu Cys Ser Pro Gly Trp Gln Gly Leu Gln Cys Glu Arg Glu Gly
    330                 335                 340

ATA CCG AGG ATG ACC CCA AAG ATA GTG GAT TTG CCA GAT CAT ATA GAA     1228
Ile Pro Arg Met Thr Pro Lys Ile Val Asp Leu Pro Asp His Ile Glu
345                 350                 355                 360

GTA AAC AGT GGT AAA TTT AAT CCC ATT TGC AAA GCT TCT GGC TGG CCG     1276
Val Asn Ser Gly Lys Phe Asn Pro Ile Cys Lys Ala Ser Gly Trp Pro
                365                 370                 375

CTA CCT ACT AAT GAA GAA ATG ACC CTG GTG AAG CCG GAT GGG ACA GTG     1324
Leu Pro Thr Asn Glu Glu Met Thr Leu Val Lys Pro Asp Gly Thr Val
            380                 385                 390

CTC CAT CCA AAA GAC TTT AAC CAT ACG GAT CAT TTC TCA GTA GCC ATA     1372
Leu His Pro Lys Asp Phe Asn His Thr Asp His Phe Ser Val Ala Ile
        395                 400                 405

TTC ACC ATC CAC CGG ATC CTC CCC CCT GAC TCA GGA GTT TGG GTC TGC     1420
Phe Thr Ile His Arg Ile Leu Pro Pro Asp Ser Gly Val Trp Val Cys
    410                 415                 420

AGT GTG AAC ACA GTG GCT GGG ATG GTG GAA AAG CCC TTC AAC ATT TCT     1468
Ser Val Asn Thr Val Ala Gly Met Val Glu Lys Pro Phe Asn Ile Ser
425                 430                 435                 440

GTT AAA GTT CTT CCA AAG CCC CTG AAT GCC CCA AAC GTG ATT GAC ACT     1516
Val Lys Val Leu Pro Lys Pro Leu Asn Ala Pro Asn Val Ile Asp Thr
                445                 450                 455

GGA CAT AAC TTT GCT GTC ATC AAC ATC AGC TCT GAG CCT TAC TTT GGG     1564
Gly His Asn Phe Ala Val Ile Asn Ile Ser Ser Glu Pro Tyr Phe Gly
            460                 465                 470
```

```
GAT GGA CCA ATC AAA TCC AAG AAG CTT CTA TAC AAA CCC GTT AAT CAC      1612
Asp Gly Pro Ile Lys Ser Lys Lys Leu Leu Tyr Lys Pro Val Asn His
        475                 480                 485

TAT GAG GCT TGG CAA CAT ATT CAA GTG ACA AAT GAG ATT GTT ACA CTC      1660
Tyr Glu Ala Trp Gln His Ile Gln Val Thr Asn Glu Ile Val Thr Leu
    490                 495                 500

AAC TAT TTG GAA CCT CGG ACA GAA TAT GAA CTC TGT GTG CAA CTG GTC      1708
Asn Tyr Leu Glu Pro Arg Thr Glu Tyr Glu Leu Cys Val Gln Leu Val
505                 510                 515                 520

CGT CGT GGA GAG GGT GGG GAA GGG CAT CCT GGA CCT GTG AGA CGC TTC      1756
Arg Arg Gly Glu Gly Gly Glu Gly His Pro Gly Pro Val Arg Arg Phe
                525                 530                 535

ACA ACA GCT TCT ATC GGA CTC CCT CCT CCA AGA GGT CTA AAT CTC CTG      1804
Thr Thr Ala Ser Ile Gly Leu Pro Pro Pro Arg Gly Leu Asn Leu Leu
            540                 545                 550

CCT AAA AGT CAG ACC ACT CTA AAT TTG ACC TGG CAA CCA ATA TTT CCA      1852
Pro Lys Ser Gln Thr Thr Leu Asn Leu Thr Trp Gln Pro Ile Phe Pro
        555                 560                 565

AGC TCG GAA GAT GAC TTT TAT GTT GAA GTG GAG AGA AGG TCT GTG CAA      1900
Ser Ser Glu Asp Asp Phe Tyr Val Glu Val Glu Arg Arg Ser Val Gln
    570                 575                 580

AAA AGT GAT CAG CAG AAT ATT AAA GTT CCA GGC AAC TTG ACT TCG GTG      1948
Lys Ser Asp Gln Gln Asn Ile Lys Val Pro Gly Asn Leu Thr Ser Val
585                 590                 595                 600

CTA CTT AAC AAC TTA CAT CCC AGG GAG CAG TAC GTG GTC CGA GCT AGA      1996
Leu Leu Asn Asn Leu His Pro Arg Glu Gln Tyr Val Val Arg Ala Arg
                605                 610                 615

GTC AAC ACC AAG GCC CAG GGG GAA TGG AGT GAA GAT CTC ACT GCT TGG      2044
Val Asn Thr Lys Ala Gln Gly Glu Trp Ser Glu Asp Leu Thr Ala Trp
            620                 625                 630

ACC CTT AGT GAC ATT CTT CCT CCT CAA CCA GAA AAC ATC AAG ATT TCC      2092
Thr Leu Ser Asp Ile Leu Pro Pro Gln Pro Glu Asn Ile Lys Ile Ser
        635                 640                 645

AAC ATT ACA CAC TCC TCG GCT GTG ATT TCT TGG ACA ATA TTG GAT GGC      2140
Asn Ile Thr His Ser Ser Ala Val Ile Ser Trp Thr Ile Leu Asp Gly
    650                 655                 660

TAT TCT ATT TCT TCT ATT ACT ATC CGT TAC AAG GTT CAA GGC AAG AAT      2188
Tyr Ser Ile Ser Ser Ile Thr Ile Arg Tyr Lys Val Gln Gly Lys Asn
665                 670                 675                 680

GAA GAC CAG CAC GTT GAT GTG AAG ATA AAG AAT GCC ACC ATC ATT CAG      2236
Glu Asp Gln His Val Asp Val Lys Ile Lys Asn Ala Thr Ile Ile Gln
                685                 690                 695

TAT CAG CTC AAG GGC CTA GAG CCT GAA ACA GCA TAC CAG GTG GAC ATT      2284
Tyr Gln Leu Lys Gly Leu Glu Pro Glu Thr Ala Tyr Gln Val Asp Ile
            700                 705                 710
```

```
TTT GCA GAG AAC AAC ATA GGG TCA AGC AAC CCA GCC TTT TCT CAT GAA      2332
Phe Ala Glu Asn Asn Ile Gly Ser Ser Asn Pro Ala Phe Ser His Glu
        715                 720                 725

CTG GTG ACC CTC CCA GAA TCT CAA GCA CCA GCG GAC CTC GGA GGG GGG      2380
Leu Val Thr Leu Pro Glu Ser Gln Ala Pro Ala Asp Leu Gly Gly Gly
    730                 735                 740

AAG ATG CTG CTT ATA GCC ATC CTT GGC TCT GCT GGA ATG ACC TGC CTG      2428
Lys Met Leu Leu Ile Ala Ile Leu Gly Ser Ala Gly Met Thr Cys Leu
745                 750                 755                 760

ACT GTG CTG TTG GCC TTT CTG ATC ATA TTG CAA TTG AAG AGG GCA AAT      2476
Thr Val Leu Leu Ala Phe Leu Ile Ile Leu Gln Leu Lys Arg Ala Asn
                765                 770                 775

GTG CAA AGG AGA ATG GCC CAA GCC TTC CAA AAC GTG AGG GAA GAA CCA      2524
Val Gln Arg Arg Met Ala Gln Ala Phe Gln Asn Val Arg Glu Glu Pro
            780                 785                 790

GCT GTG CAG TTC AAC TCA GGG ACT CTG GCC CTA AAC AGG AAG GTC AAA      2572
Ala Val Gln Phe Asn Ser Gly Thr Leu Ala Leu Asn Arg Lys Val Lys
        795                 800                 805

AAC AAC CCA GAT CCT ACA ATT TAT CCA GTG CTT GAC TGG AAT GAC ATC      2620
Asn Asn Pro Asp Pro Thr Ile Tyr Pro Val Leu Asp Trp Asn Asp Ile
    810                 815                 820

AAA TTT CAA GAT GTG ATT GGG GAG GGC AAT TTT GGC CAA GTT CTT AAG      2668
Lys Phe Gln Asp Val Ile Gly Glu Gly Asn Phe Gly Gln Val Leu Lys
825                 830                 835                 840

GCG CGC ATC AAG AAG GAT GGG TTA CGG ATG GAT GCT GCC ATC AAA AGA      2716
Ala Arg Ile Lys Lys Asp Gly Leu Arg Met Asp Ala Ala Ile Lys Arg
                845                 850                 855

ATG AAA GAA TAT GCC TCC AAA GAT GAT CAC AGG GAC TTT GCA GGA GAA      2764
Met Lys Glu Tyr Ala Ser Lys Asp Asp His Arg Asp Phe Ala Gly Glu
            860                 865                 870

CTG GAA GTT CTT TGT AAA CTT GGA CAC CAT CCA AAC ATC ATC AAT CTC      2812
Leu Glu Val Leu Cys Lys Leu Gly His His Pro Asn Ile Ile Asn Leu
        875                 880                 885

TTA GGA GCA TGT GAA CAT CGA GGC TAC TTG TAC CTG GCC ATT GAG TAC      2860
Leu Gly Ala Cys Glu His Arg Gly Tyr Leu Tyr Leu Ala Ile Glu Tyr
    890                 895                 900

GCG CCC CAT GGA AAC CTT CTG GAC TTC CTT CGC AAG AGC CGT GTG CTG      2908
Ala Pro His Gly Asn Leu Leu Asp Phe Leu Arg Lys Ser Arg Val Leu
905                 910                 915                 920

GAG ACG GAC CCA GCA TTT GCC ATT GCC AAT AGC ACC GCG TCC ACA CTG      2956
Glu Thr Asp Pro Ala Phe Ala Ile Ala Asn Ser Thr Ala Ser Thr Leu
                925                 930                 935

TCC TCC CAG CAG CTC CTT CAC TTC GCT GCC GAC GTG GCC CGG GGC ATG      3004
Ser Ser Gln Gln Leu Leu His Phe Ala Ala Asp Val Ala Arg Gly Met
            940                 945                 950
```

```
GAC TAC TTG AGC CAA AAA CAG TTT ATC CAC AGG GAT CTG GCT GCC AGA      3052
Asp Tyr Leu Ser Gln Lys Gln Phe Ile His Arg Asp Leu Ala Ala Arg
        955                 960                 965

AAC ATT TTA GTT GGT GAA AAC TAT GTG GCA AAA ATA GCA GAT TTT GGA      3100
Asn Ile Leu Val Gly Glu Asn Tyr Val Ala Lys Ile Ala Asp Phe Gly
    970                 975                 980

TTG TCC CGA GGT CAA GAG GTG TAC GTG AAA AAG ACA ATG GGA AGG CTC      3148
Leu Ser Arg Gly Gln Glu Val Tyr Val Lys Lys Thr Met Gly Arg Leu
985                 990                 995                 1000

CCA GTG CGC TGG ATG GCC ATC GAG TCA CTG AAT TAC AGT GTG TAC ACA      3196
Pro Val Arg Trp Met Ala Ile Glu Ser Leu Asn Tyr Ser Val Tyr Thr
                1005                1010                1015

ACC AAC AGT GAT GTA TGG TCC TAT GGT GTG TTA CTA TGG GAG ATT GTT      3244
Thr Asn Ser Asp Val Trp Ser Tyr Gly Val Leu Leu Trp Glu Ile Val
            1020                1025                1030

AGC TTA GGA GGC ACA CCC TAC TGC GGG ATG ACT TGT GCA GAA CTC TAC      3292
Ser Leu Gly Gly Thr Pro Tyr Cys Gly Met Thr Cys Ala Glu Leu Tyr
        1035                1040                1045

GAG AAG CTG CCC CAG GGC TAC AGA CTG GAG AAG CCC CTG AAC TGT GAT      3340
Glu Lys Leu Pro Gln Gly Tyr Arg Leu Glu Lys Pro Leu Asn Cys Asp
    1050                1055                1060

GAT GAG GTG TAT GAT CTA ATG AGA CAA TGC TGG CGG GAG AAG CCT TAT      3388
Asp Glu Val Tyr Asp Leu Met Arg Gln Cys Trp Arg Glu Lys Pro Tyr
1065                1070                1075                1080

GAG AGG CCA TCA TTT GCC CAG ATA TTG GTG TCC TTA AAC AGA ATG TTA      3436
Glu Arg Pro Ser Phe Ala Gln Ile Leu Val Ser Leu Asn Arg Met Leu
                1085                1090                1095

GAG GAG CGA AAG ACC TAC GTG AAT ACC ACG CTT TAT GAG AAG TTT ACT      3484
Glu Glu Arg Lys Thr Tyr Val Asn Thr Thr Leu Tyr Glu Lys Phe Thr
            1100                1105                1110

TAT GCA GGA ATT GAC TGT TCT GCT GAA GAA GCG GCC TAGGACAGAA           3530
Tyr Ala Gly Ile Asp Cys Ser Ala Glu Glu Ala Ala
        1115                1120                112

CATCTGTATA CCCTCTGTTT CCCTTTCACT GGCATGGGAG ACCCTTGACA ACTGCTGAGA   3590

AAACATGCCT CTGCCAAAGG ATGTGATATA TAAGTGTACA TATGTGCTGG AATTCTAACA   3650

AGTCATAGGT TAATATTTAA GACACTGAAA AATCTAAGTG ATATAAATCA GATTCTTCTC   3710

TCTCATTTTA TCCCTCACCT GTAGCATGCC AGTCCCGTTT CATTTAGTCA TGTGACCACT   3770

CTGTCTTGTG TTTCCACAGC CTGCAAGTTC AGTCCAGGAT GCTAACATCT AAAAATAGAC   3830

TTAAATCTCA TTGCTTACAA GCCTAAGAAT CTTTAGAGAA GTATACATAA GTTTAGGATA   3890

AAATAATGGG ATTTTCTTTT CTTTTCTCTG GTAATATTGA CTTGTATATT TTAAGAAATA   3950

ACAGAAAGCC TGGGTGACAT TTGGGAGACA TGTGACATTT ATATATTGAA TTAATATCCC   4010

TACATGTATT GCACATTGTA AAAAGTTTTA GTTTTGATGA GTTGTGAGTT TACCTTGTAT   4070
```

ACTGTAGGCA CACTTTGCAC TGATATATCA TGAGTGAATA AATGTCTTGC CTACTCAAAA 4130

AAAAAAAA 4138

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1124 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asp Ser Leu Ala Ser Leu Val Leu Cys Gly Val Ser Leu Leu Leu
1 5 10 15

Ser Gly Thr Val Glu Gly Ala Met Asp Leu Ile Leu Ile Asn Ser Leu
20 25 30

Pro Leu Val Ser Asp Ala Glu Thr Ser Leu Thr Cys Ile Ala Ser Gly
35 40 45

Trp Arg Pro His Glu Pro Ile Thr Ile Gly Arg Asp Phe Glu Ala Leu
50 55 60

Met Asn Gln His Gln Asp Pro Leu Glu Val Thr Gln Asp Val Thr Arg
65 70 75 80

Glu Trp Ala Lys Lys Val Val Trp Lys Arg Glu Lys Ala Ser Lys Ile
85 90 95

Asn Gly Ala Tyr Phe Cys Glu Gly Arg Val Arg Gly Glu Ala Ile Arg
100 105 110

Ile Arg Thr Met Lys Met Arg Gln Gln Ala Ser Phe Leu Pro Ala Thr
115 120 125

Leu Thr Met Thr Val Asp Lys Gly Asp Asn Val Asn Ile Ser Phe Lys
130 135 140

Lys Val Leu Ile Lys Glu Glu Asp Ala Val Ile Tyr Lys Asn Gly Ser
145 150 155 160

Phe Ile His Ser Val Pro Arg His Glu Val Pro Asp Ile Leu Glu Val
165 170 175

His Leu Pro His Ala Gln Pro Gln Asp Ala Gly Val Tyr Ser Ala Arg
180 185 190

Tyr Ile Gly Gly Asn Leu Phe Thr Ser Ala Phe Thr Arg Leu Ile Val
195 200 205

Arg Arg Cys Glu Ala Gln Lys Trp Gly Pro Glu Cys Asn His Leu Cys
210 215 220

Thr Ala Cys Met Asn Asn Gly Val Cys His Glu Asp Thr Gly Glu Cys
225 230 235 240

Ile Cys Pro Pro Gly Phe Met Gly Arg Thr Cys Glu Lys Ala Cys Glu
245 250 255

Leu His Thr Phe Gly Arg Thr Cys Lys Glu Arg Cys Ser Gly Gln Glu
260 265 270

Gly Cys Lys Ser Tyr Val Phe Cys Leu Pro Asp Pro Tyr Gly Cys Ser
275 280 285

Cys Ala Thr Gly Trp Lys Gly Leu Gln Cys Asn Glu Ala Cys His Pro
290 295 300

Gly Phe Tyr Gly Pro Asp Cys Lys Leu Arg Cys Ser Cys Asn Asn Gly
305 310 315 320

Glu Met Cys Asp Arg Phe Gln Gly Cys Leu Cys Ser Pro Gly Trp Gln
325 330 335

Gly Leu Gln Cys Glu Arg Glu Gly Ile Pro Arg Met Thr Pro Lys Ile
340 345 350

Val Asp Leu Pro Asp His Ile Glu Val Asn Ser Gly Lys Phe Asn Pro
355 360 365

Ile Cys Lys Ala Ser Gly Trp Pro Leu Pro Thr Asn Glu Glu Met Thr
370 375 380

Leu Val Lys Pro Asp Gly Thr Val Leu His Pro Lys Asp Phe Asn His
385 390 395 400

Thr Asp His Phe Ser Val Ala Ile Phe Thr Ile His Arg Ile Leu Pro
405 410 415

Pro Asp Ser Gly Val Trp Val Cys Ser Val Asn Thr Val Ala Gly Met
420 425 430

Val Glu Lys Pro Phe Asn Ile Ser Val Lys Val Leu Pro Lys Pro Leu
435 440 445

Asn Ala Pro Asn Val Ile Asp Thr Gly His Asn Phe Ala Val Ile Asn
450 455 460

Ile Ser Ser Glu Pro Tyr Phe Gly Asp Gly Pro Ile Lys Ser Lys Lys
465 470 475 480

Leu Leu Tyr Lys Pro Val Asn His Tyr Glu Ala Trp Gln His Ile Gln
485 490 495

Val Thr Asn Glu Ile Val Thr Leu Asn Tyr Leu Glu Pro Arg Thr Glu
500 505 510

Tyr Glu Leu Cys Val Gln Leu Val Arg Arg Gly Glu Gly Gly Glu Gly
515 520 525

His Pro Gly Pro Val Arg Arg Phe Thr Thr Ala Ser Ile Gly Leu Pro
530 535 540

Pro Pro Arg Gly Leu Asn Leu Leu Pro Lys Ser Gln Thr Thr Leu Asn
545 550 555 560

Leu Thr Trp Gln Pro Ile Phe Pro Ser Ser Glu Asp Asp Phe Tyr Val
565 570 575

Glu Val Glu Arg Arg Ser Val Gln Lys Ser Asp Gln Gln Asn Ile Lys
580 585 590

Val Pro Gly Asn Leu Thr Ser Val Leu Leu Asn Asn Leu His Pro Arg
595 600 605

Glu Gln Tyr Val Val Arg Ala Arg Val Asn Thr Lys Ala Gln Gly Glu
610 615 620

Trp Ser Glu Asp Leu Thr Ala Trp Thr Leu Ser Asp Ile Leu Pro Pro
625 630 635 640

Gln Pro Glu Asn Ile Lys Ile Ser Asn Ile Thr His Ser Ser Ala Val
645 650 655

Ile Ser Trp Thr Ile Leu Asp Gly Tyr Ser Ile Ser Ser Ile Thr Ile
660 665 670

Arg Tyr Lys Val Gln Gly Lys Asn Glu Asp Gln His Val Asp Val Lys
675 680 685

Ile Lys Asn Ala Thr Ile Ile Gln Tyr Gln Leu Lys Gly Leu Glu Pro
690 695 700

Glu Thr Ala Tyr Gln Val Asp Ile Phe Ala Glu Asn Asn Ile Gly Ser
705 710 715 720

Ser Asn Pro Ala Phe Ser His Glu Leu Val Thr Leu Pro Glu Ser Gln
725 730 735

Ala Pro Ala Asp Leu Gly Gly Gly Lys Met Leu Leu Ile Ala Ile Leu
740 745 750

Gly Ser Ala Gly Met Thr Cys Leu Thr Val Leu Leu Ala Phe Leu Ile
755 760 765

Ile Leu Gln Leu Lys Arg Ala Asn Val Gln Arg Arg Met Ala Gln Ala
770 775 780

Phe Gln Asn Val Arg Glu Glu Pro Ala Val Gln Phe Asn Ser Gly Thr
785 790 795 800

Leu Ala Leu Asn Arg Lys Val Lys Asn Asn Pro Asp Pro Thr Ile Tyr
805 810 815

Pro Val Leu Asp Trp Asn Asp Ile Lys Phe Gln Asp Val Ile Gly Glu
820 825 830

Gly Asn Phe Gly Gln Val Leu Lys Ala Arg Ile Lys Lys Asp Gly Leu
835 840 845

Arg Met Asp Ala Ala Ile Lys Arg Met Lys Glu Tyr Ala Ser Lys Asp
850 855 860

Asp His Arg Asp Phe Ala Gly Glu Leu Glu Val Leu Cys Lys Leu Gly
865 870 875 880

His His Pro Asn Ile Ile Asn Leu Leu Gly Ala Cys Glu His Arg Gly
885 890 895

Tyr Leu Tyr Leu Ala Ile Glu Tyr Ala Pro His Gly Asn Leu Leu Asp
900 905 910

Phe Leu Arg Lys Ser Arg Val Leu Glu Thr Asp Pro Ala Phe Ala Ile
915 920 925

Ala Asn Ser Thr Ala Ser Thr Leu Ser Ser Gln Gln Leu Leu His Phe
930 935 940

Ala Ala Asp Val Ala Arg Gly Met Asp Tyr Leu Ser Gln Lys Gln Phe
945 950 955 960

Ile His Arg Asp Leu Ala Ala Arg Asn Ile Leu Val Gly Glu Asn Tyr
965 970 975

Val Ala Lys Ile Ala Asp Phe Gly Leu Ser Arg Gly Gln Glu Val Tyr
980 985 990

Val Lys Lys Thr Met Gly Arg Leu Pro Val Arg Trp Met Ala Ile Glu
995 1000 1005

Ser Leu Asn Tyr Ser Val Tyr Thr Thr Asn Ser Asp Val Trp Ser Tyr
1010 1015 1020

Gly Val Leu Leu Trp Glu Ile Val Ser Leu Gly Gly Thr Pro Tyr Cys
1025 1030 1035 1040

Gly Met Thr Cys Ala Glu Leu Tyr Glu Lys Leu Pro Gln Gly Tyr Arg
1045 1050 1055

Leu Glu Lys Pro Leu Asn Cys Asp Asp Glu Val Tyr Asp Leu Met Arg
1060 1065 1070

Gln Cys Trp Arg Glu Lys Pro Tyr Glu Arg Pro Ser Phe Ala Gln Ile
1075 1080 1085

Leu Val Ser Leu Asn Arg Met Leu Glu Glu Arg Lys Thr Tyr Val Asn
1090 1095 1100

Thr Thr Leu Tyr Glu Lys Phe Thr Tyr Ala Gly Ile Asp Cys Ser Ala
1105 1110 1115 1120

Glu Glu Ala Ala

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 239 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Val Pro Trp Leu Gly Ala Cys Glu His Arg Gly Tyr Leu Tyr Leu Ala
1               5                  10                  15

Ile Glu Tyr Ala Pro His Gly Asn Leu Leu Asp Phe Leu Arg Lys Ser
            20                  25                  30

Arg Val Leu Glu Thr Asp Pro Ala Phe Ala Ile Ala Asn Ser Thr Ala
        35                  40                  45

Ser Ile Met Ser Ser Gln Gln Leu Leu His Phe Ala Ala Asp Val Ala
    50                  55                  60

Arg Gly Met Asp Tyr Leu Ser Gln Lys Gln Phe Ile His Arg Asp Leu
65                  70                  75                  80

Ala Ala Arg Asn Ile Leu Val Gly Glu Asn Tyr Ile Ala Lys Ile Ala
                85                  90                  95

Asp Phe Gly Leu Ser Arg Gly Gln Glu Val Tyr Val Lys Lys Thr Met
            100                 105                 110

Gly Arg Leu Pro Val Arg Trp Met Ala Ile Glu Ser Leu Asn Tyr Ser
        115                 120                 125

Val Tyr Thr Thr Asn Ser Asp Val Trp Ser Tyr Gly Val Leu Leu Trp
    130                 135                 140

Glu Ile Val Ser Leu Gly Gly Thr Pro Tyr Cys Gly Met Thr Cys Ala
145                 150                 155                 160

Glu Leu Tyr Glu Lys Leu Pro Gln Gly Tyr Arg Leu Glu Lys Pro Leu
                165                 170                 175

Asn Cys Asp Asp Glu Val Tyr Asp Leu Met Arg Gln Cys Trp Arg Glu
            180                 185                 190

Lys Pro Tyr Glu Arg Pro Ser Phe Ala Gln Ile Leu Val Ser Leu Asn
        195                 200                 205

Arg Met Leu Glu Glu Arg Lys Thr Tyr Val Asn Thr Thr Leu Tyr Glu
    210                 215                 220

Lys Phe Thr Tyr Ala Gly Ile Asp Cys Ser Ala Glu Glu Ala Ala
225                 230                 235
```

CLAIMS

What is claimed is:

1. An isolated DNA sequence encoding a biologically active human *ork* polypeptide, wherein the DNA sequence is selected from the group consisting of:

(a) DNA comprising the coding region of the DNA sequence of SEQ ID NO:1;

(b) DNA sequences that hybridize under moderately stringent conditions to the DNA of (a) and which encode a biologically active *ork*; and

(c) DNA sequences that are degenerate as a result of the genetic code to a DNA sequence of (a) or (b), and which encode a biologically active *ork*.

2. An isolated DNA sequence according to claim 1 wherein the *ork* polypeptide is a soluble human *ork* polypeptide.

3. An isolated DNA sequence according to claim 2, wherein said soluble human *ork* polypeptide comprises amino acids 19-745 of the amino acid sequence of SEQ ID NO:1.

4. An isolated DNA sequence comprising a nucleotide sequence selected from the group consisting of nucleotides 149-3520, 203-3520, 149-2383, and 203-2383 of SEQ ID NO:1.

5. An expression vector comprising a DNA sequence according to claim 1.

6. An expression vector comprising a DNA sequence according to claim 2.

7. An expression vector comprising a DNA sequence according to claim 3.

8. An expression vector comprising a DNA sequence according to claim 4.

9. A process for preparing an *ork* polypeptide, comprising culturing a host cell transformed with an expression vector according to claim 5 under conditions promoting expression of *ork*, and recovering the *ork* polypeptide from the culture.

10. A process for preparing an *ork* polypeptide, comprising culturing a host cell transformed with an expression vector according to claim 6 under conditions promoting expression of *ork* and recovering the *ork* polypeptide from the culture.

11. A process for preparing an *ork* polypeptide, comprising culturing a host cell transformed with an expression vectoraccording to claim 7 under conditions promoting expression of *ork*, and recovering the *ork* polypeptide from the culture.

12. A process for preparing an *ork* polypeptide, comprising culturing a host cell transformed with an expression vector according to claim 8 under conditions promoting expression of *ork*, and recovering the *ork* polypeptide from the culture.

13. A substantially homogeneous purified biologically active *ork* protein encoded by a DNA sequence selected from the group consisting of:

(a) DNA comprising the coding region of the DNA sequence of SEQ ID NO:1;

(b) DNA sequences that hybridize under moderately stringent conditions to the DNA of (a) and which encode a biologically active *ork*; and

(c) DNA sequences that are degenerate as a result of the genetic code to a DNA sequence of (a) or (b), and which encode a biologically active *ork*.

14. A purified *ork* according to claim 13, wherein said *ork* is human *ork*.

15. A purified *ork* according to claim 14, wherein said *ork* is a soluble human *ork*.

16. A purified human *ork* protein comprising an amino acid sequence selected from the group consisting of amino acids 19-1124 of SEQ ID NO:2 and amino acids 19-745 of SEQ ID NO:2.

17. A purified *ork* protein comprising an extracellular domain that comprises (from N- to C-terminus) an immunoglobulin-like domain, three EGF-like cysteine repeats, a sequence containing a pair of cysteine residues, and three FNIII-like repeats, wherein the amino acid sequence of said extracellular domain is at least 80% identical to the sequence presented as amino acids 19-745 of SEQ ID NO:1.

18. A purified *ork* protein according to claim 17, wherein said protein additionally contains a transmembrane region and a cytoplasmic domain.

19. An antibody immunoreactive with *ork* or an *ork* immunogen.

20. An antisense or sense oligonucleotide that can inhibit transcription or translation of *ork*, comprising a sequence of at least about 14 nucleotides of the DNA sequence of SEQ ID NO:1, or its DNA or RNA complement.

FIGURE 1

```
                                                CTTCTGTGCTGTTCCTTCTTGCCTCTAA    28
          CTTGTAAACAAGACGTACTAGGACGATGCTAATGGAAAGTCACAAACC                      76
          GCTGGGTTTTTGAAAGGATCCTTGGGACCTCATGCACATTTGTGGAAA                      124
          CTGGATGGAGAGATTTGGGGAAGCATGGACTCTTTAGCCAGCTTAGTT                      172
1                                                 M  D  S  L  A  S  L  V
          CTCTGTGGAGTCAGCTTGCTCCTTTCTGGAACTGTGGAAGGTGCCATG                      220
9         L  C  G  V  S  L  L  L  S  G  T  V  E  G  A  M
          GACTTGATCTTGATCAATTCCCTACCTCTTGTATCTGATGCTGAAACA                      268
25        D  L  I  L  I  N  S  L  P  L  V  S  D  A  E  T
          TCTCTCACCTGCATTGCCTCTGGGTGGCGCCCCCATGAGCCCATCACC                      316
41        S  L  T  C  I  A  S  G  W  R  P  H  E  P  I  T
          ATAGGAAGGGACTTTGAAGCCTTAATGAACCAGCACCAGGATCCGCTG                      364
57        I  G  R  D  F  E  A  L  M  N  Q  H  Q  D  P  L
          GAAGTTACTCAAGATGTGACCAGAGAATGGGCTAAAAAAGTTGTTTGG                      412
73        E  V  T  Q  D  V  T  R  E  W  A  K  K  V  V  W
          AAGAGAGAAAAGGCTAGTAAGATCAATGGTGCTTATTTCTGTGAAGGG                      460
89        K  R  E  K  A  S  K  I  N  G  A  Y  F  C  E  G
```

FIGURE 1, CONTINUED

```
     CGAGTTCGAGGAGAGGCAATCAGGATACGAACCATGAAGATGCGTCAA   508
105   R  V  R  G  E  A  I  R  I  R  T  M  K  M  R  Q
     CAAGCTTCCTTCCTACCAGCTACTTTAACTATGACTGTGGACAAGGGA   556
121   Q  A  S  F  L  P  A  T  L  T  M  T  V  D  K  G
     GATAACGTGAACATATCTTTCAAAAAGGTATTGATTAAAGAAGAAGAT   604
137   D  N  V  N  I  S  F  K  K  V  L  I  K  E  E  D
     GCAGTGATTTACAAAAATGGTTCCTTCATCCATTCAGTGCCCCGGCAT   652
153   A  V  I  Y  K  N  G  S  F  I  H  S  V  P  R  H
     GAAGTACCTGATATTCTAGAAGTACACCTGCCTCATGCTCAGCCCCAG   700
169   E  V  P  D  I  L  E  V  H  L  P  H  A  Q  P  Q
     GATGCTGGAGTGTACTCGGCCAGGTATATAGGAGGAAACCTCTTCACC   748
185   D  A  G  V  Y  S  A  R  Y  I  G  G  N  L  F  T
     TCGGCCTTCACCAGGCTGATAGTCCGGAGATGTGAAGCCCAGAAGTGG   796
201   S  A  F  T  R  L  I  V  R  R  C  E  A  Q  K  W
     GGACCTGAATGCAACCATCTCTGTACTGCTTGTATGAACAATGGTGTC   844
217   G  P  E  C  N  H  L  C  T  A  C  M  N  N  G  V
     TGCCATGAAGATACTGGAGAATGCATTTGCCCTCCTGGGTTTATGGGA   892
233   C  H  E  D  T  G  E  C  I  C  P  P  G  F  M  G
     AGGACGTGTGAGAAGGCTTGTGAACTGCACACGTTTGGCAGAACTTGT   940
249   R  T  C  E  K  A  C  E  L  H  T  F  G  R  T  C
```

FIGURE 1, CONTINUED

```
                AAAGAAAGGTGCAGTGGACAAGAGGGATGCAAGTCTTATGTGTTCTGT   988
265              K   E   R   C   S   G   Q   E   G   C   K   S   Y   V   F   C

                CTCCCTGACCCCTATGGGTGTTCCTGTGCCACAGGCTGGAAGGGTCTG  1036
281              L   P   D   P   Y   G   C   S   C   A   T   G   W   K   G   L

                CAGTGCAATGAAGCATGCCACCCTGGTTTTTACGGGCCAGATTGTAAG  1084
297              Q   C   N   E   A   C   H   P   G   F   Y   G   P   D   C   K

                CTTAGGTGCAGCTGCAACAATGGGGAGATGTGTGATCGCTTCCAAGGA  1132
313              L   R   C   S   C   N   N   G   E   M   C   D   R   F   Q   G

                TGTCTCTGCTCTCCAGGATGGCAGGGGCTCCAGTGTGAGAGAGAAGGC  1180
329              C   L   C   S   P   G   W   Q   G   L   Q   C   E   R   E   G

                ATACCGAGGATGACCCCAAAGATAGTGGATTTGCCAGATCATATAGAA  1228
345              I   P   R   M   T   P   K   I   V   D   L   P   D   H   I   E

                GTAAACAGTGGTAAATTTAATCCCATTTGCAAAGCTTCTGGCTGGCCG  1276
361              V   N   S   G   K   F   N   P   I   C   K   A   S   G   W   P

                CTACCTACTAATGAAGAAATGACCCTGGTGAAGCCGGATGGGACAGTG  1324
377              L   P   T   N   E   E   M   T   L   V   K   P   D   G   T   V

                CTCCATCCAAAAGACTTTAACCATACGGATCATTTCTCAGTAGCCATA  1372
393              L   H   P   K   D   F   N   H   T   D   H   F   S   V   A   I

                TTCACCATCCACCGGATCCTCCCCCCTGACTCAGGAGTTTGGGTCTGC  1420
409              F   T   I   H   R   I   L   P   P   D   S   G   V   W   V   C
```

FIGURE 1, CONTINUED

```
     AGTGTGAACACAGTGGCTGGGATGGTGGAAAAGCCCTTCAACATTTCT  1468
425   S  V  N  T  V  A  G  M  V  E  K  P  F  N  I  S
     GTTAAAGTTCTTCCAAAGCCCCTGAATGCCCCAAACGTGATTGACACT  1516
441   V  K  V  L  P  K  P  L  N  A  P  N  V  I  D  T
     GGACATAACTTTGCTGTCATCAACATCAGCTCTGAGCCTTACTTTGGG  1564
457   G  H  N  F  A  V  I  N  I  S  S  E  P  Y  F  G
     GATGGACCAATCAAATCCAAGAAGCTTCTATACAAACCCGTTAATCAC  1612
473   D  G  P  I  K  S  K  K  L  L  Y  K  P  V  N  H
     TATGAGGCTTGGCAACATATTCAAGTGACAAATGAGATTGTTACACTC  1660
489   Y  E  A  W  Q  H  I  Q  V  T  N  E  I  V  T  L
     AACTATTTGGAACCTCGGACAGAATATGAACTCTGTGTGCAACTGGTC  1708
505   N  Y  L  E  P  R  T  E  Y  E  L  C  V  Q  L  V
     CGTCGTGGAGAGGGTGGGGAAGGGCATCCTGGACCTGTGAGACGCTTC  1756
521   R  R  G  E  G  G  E  G  H  P  G  P  V  R  R  F
     ACAACAGCTTCTATCGGACTCCCTCCTCCAAGAGGTCTAAATCTCCTG  1804
537   T  T  A  S  I  G  L  P  P  P  R  G  L  N  L  L
     CCTAAAAGTCAGACCACTCTAAATTTGACCTGGCAACCAATATTTCCA  1852
553   P  K  S  Q  T  T  L  N  L  T  W  Q  P  I  F  P
     AGCTCGGAAGATGACTTTTATGTTGAAGTGGAGAGAAGGTCTGTGCAA  1900
569   S  S  E  D  D  F  Y  V  E  V  E  R  R  S  V  Q
```

FIGURE 1, CONTINUED

```
        AAAAGTGATCAGCAGAATATTAAAGTTCCAGGCAACTTGACTTCGGTG   1948
585      K   S   D   Q   Q   N   I   K   V   P   G   N   L   T   S   V
        CTACTTAACAACTTACATCCCAGGGAGCAGTACGTGGTCCGAGCTAGA   1996
601      L   L   N   N   L   H   P   R   E   Q   Y   V   V   R   A   R
        GTCAACACCAAGGCCCAGGGGGAATGGAGTGAAGATCTCACTGCTTGG   2044
617      V   N   T   K   A   Q   G   E   W   S   E   D   L   T   A   W
        ACCCTTAGTGACATTCTTCCTCCTCAACCAGAAAACATCAAGATTTCC   2092
633      T   L   S   D   I   L   P   P   Q   P   E   N   I   K   I   S
        AACATTACACACTCCTCGGCTGTGATTTCTTGGACAATATTGGATGGC   2140
649      N   I   T   H   S   S   A   V   I   S   W   T   I   L   D   G
        TATTCTATTTCTTCTATTACTATCCGTTACAAGGTTCAAGGCAAGAAT   2188
665      Y   S   I   S   S   I   T   I   R   Y   K   V   Q   G   K   N
        GAAGACCAGCACGTTGATGTGAAGATAAAGAATGCCACCATCATTCAG   2236
681      E   D   Q   H   V   D   V   K   I   K   N   A   T   I   I   Q
        TATCAGCTCAAGGGCCTAGAGCCTGAAACAGCATACCAGGTGGACATT   2284
697      Y   Q   L   K   G   L   E   P   E   T   A   Y   Q   V   D   I
        TTTGCAGAGAACAACATAGGGTCAAGCAACCCAGCCTTTTCTCATGAA   2332
713      F   A   E   N   N   I   G   S   S   N   P   A   F   S   H   E
        CTGGTGACCCTCCCAGAATCTCAAGCACCAGCGGACCTCGGAGGGGGG   2380
729      L   V   T   L   P   E   S   Q   A   P   A   D   L   G   G   G
```

FIGURE 1, CONTINUED

```
     AAGATGCTGCTTATAGCCATCCTTGGCTCTGCTGGAATGACCTGCCTG  2428
745    K  M  L  L  I  A  I  L  G  S  A  G  M  T  C  L
     ACTGTGCTGTTGGCCTTTCTGATCATATTGCAATTGAAGAGGGCAAAT  2476
761    T  V  L  L  A  F  L  I  I  L  Q  L  K  R  A  N
     GTGCAAAGGAGAATGGCCCAAGCCTTCCAAAACGTGAGGGAAGAACCA  2524
777    V  Q  R  R  M  A  Q  A  F  Q  N  V  R  E  E  P
     GCTGTGCAGTTCAACTCAGGGACTCTGGCCCTAAACAGGAAGGTCAAA  2572
793    A  V  Q  F  N  S  G  T  L  A  L  N  R  K  V  K
     AACAACCCAGATCCTACAATTTATCCAGTGCTTGACTGGAATGACATC  2620
809    N  N  P  D  P  T  I  Y  P  V  L  D  W  N  D  I
     AAATTTCAAGATGTGATTGGGGAGGGCAATTTTGGCCAAGTTCTTAAG  2668
825    K  F  Q  D  V  I  G  E  G  N  F  G  Q  V  L  K
     GCGCGCATCAAGAAGGATGGGTTACGGATGGATGCTGCCATCAAAAGA  2716
841    A  R  I  K  K  D  G  L  R  M  D  A  A  I  K  R
     ATGAAAGAATATGCCTCCAAAGATGATCACAGGGACTTTGCAGGAGAA  2764
857    M  K  E  Y  A  S  K  D  D  H  R  D  F  A  G  E
     CTGGAAGTTCTTTGTAAACTTGGACACCATCCAAACATCATCAATCTC  2812
873    L  E  V  L  C  K  L  G  H  H  P  N  I  I  N  L
     TTAGGAGCATGTGAACATCGAGGCTACTTGTACCTGGCCATTGAGTAC  2860
889    L  G  A  C  E  H  R  G  Y  L  Y  L  A  I  E  Y
```

FIGURE 1, CONTINUED

```
       GCGCCCCATGGAAACCTTCTGGACTTCCTTCGCAAGAGCCGTGTGCTG   2908
905     A  P  H  G  N  L  L  D  F  L  R  K  S  R  V  L
       GAGACGGACCCAGCATTTGCCATTGCCAATAGCACCGCGTCCACACTG   2956
921     E  T  D  P  A  F  A  I  A  N  S  T  A  S  T  L
       TCCTCCCAGCAGCTCCTTCACTTCGCTGCCGACGTGGCCCGGGGCATG   3004
937     S  S  Q  Q  L  L  H  F  A  A  D  V  A  R  G  M
       GACTACTTGAGCCAAAAACAGTTTATCCACAGGGATCTGGCTGCCAGA   3052
953     D  Y  L  S  Q  K  Q  F  I  H  R  D  L  A  A  R
       AACATTTTAGTTGGTGAAAACTATGTGGCAAAAATAGCAGATTTTGGA   3100
969     N  I  L  V  G  E  N  Y  V  A  K  I  A  D  F  G
       TTGTCCCGAGGTCAAGAGGTGTACGTGAAAAAGACAATGGGAAGGCTC   3148
985     L  S  R  G  Q  E  V  Y  V  K  K  T  M  G  R  L
       CCAGTGCGCTGGATGGCCATCGAGTCACTGAATTACAGTGTGTACACA   3196
1001    P  V  R  W  M  A  I  E  S  L  N  Y  S  V  Y  T
       ACCAACAGTGATGTATGGTCCTATGGTGTGTTACTATGGGAGATTGTT   3244
1017    T  N  S  D  V  W  S  Y  G  V  L  L  W  E  I  V
       AGCTTAGGAGGCACACCCTACTGCGGGATGACTTGTGCAGAACTCTAC   3292
1033    S  L  G  G  T  P  Y  C  G  M  T  C  A  E  L  Y
       GAGAAGCTGCCCCAGGGCTACAGACTGGAGAAGCCCCTGAACTGTGAT   3340
1049    E  K  L  P  Q  G  Y  R  L  E  K  P  L  N  C  D
```

FIGURE 1, CONTINUED

```
     GATGAGGTGTATGATCTAATGAGACAATGCTGGCGGGAGAAGCCTTAT   3388
1065  D  E  V  Y  D  L  M  R  Q  C  W  R  E  K  P  Y

     GAGAGGCCATCATTTGCCCAGATATTGGTGTCCTTAAACAGAATGTTA   3436
1081  E  R  P  S  F  A  Q  I  L  V  S  L  N  R  M  L

     GAGGAGCGAAAGACCTACGTGAATACCACGCTTTATGAGAAGTTTACT   3484
1097  E  E  R  K  T  Y  V  N  T  T  L  Y  E  K  F  T

     TATGCAGGAATTGACTGTTCTGCTGAAGAAGCGGCCTAGGACAGAACA   3532
1113  Y  A  G  I  D  C  S  A  E  E  A  A  *

     TCTGTATACCCTCTGTTTCCCTTTCACTGGCATGGGAGACCCTTGACA   3580

     ACTGCTGAGAAAACATGCCTCTGCCAAAGGATGTGATATATAAGTGTA   3628

     CATATGTGCTGGAATTCTAACAAGTCATAGGTTAATATTTAAGACACT   3676

     GAAAAATCTAAGTGATATAAATCAGATTCTTCTCTCTCATTTTATCCC   3724

     TCACCTGTAGCATGCCAGTCCCGTTTCATTTAGTCATGTGACCACTCT   3772

     GTCTTGTGTTTCCACAGCCTGCAAGTTCAGTCCAGGATGCTAACATCT   3820

     AAAAATAGACTTAAATCTCATTGCTTACAAGCCTAAGAATCTTTAGAG   3868

     AAGTATACATAAGTTTAGGATAAAATAATGGGATTTTCTTTTCTTTTC   3916

     TCTGGTAATATTGACTTGTATATTTTAAGAAATAACAGAAAGCCTGGG   3964

     TGACATTTGGGAGACATGTGACATTTATATATTGAATTAATATCCCTA   4012
```

FIGURE 1, CONTINUED

```
CATGTATTGCACATTGTAAAAAGTTTTAGTTTTGATGAGTTGTGAGTT  4060
TACCTTGTATACTGTAGGCACACACTTTGCACTGATATATCATGAGTGAA  4108
TAAATGTCTTGCCTACTCAAAAAAAAAAAA  4138
```

Figure 2

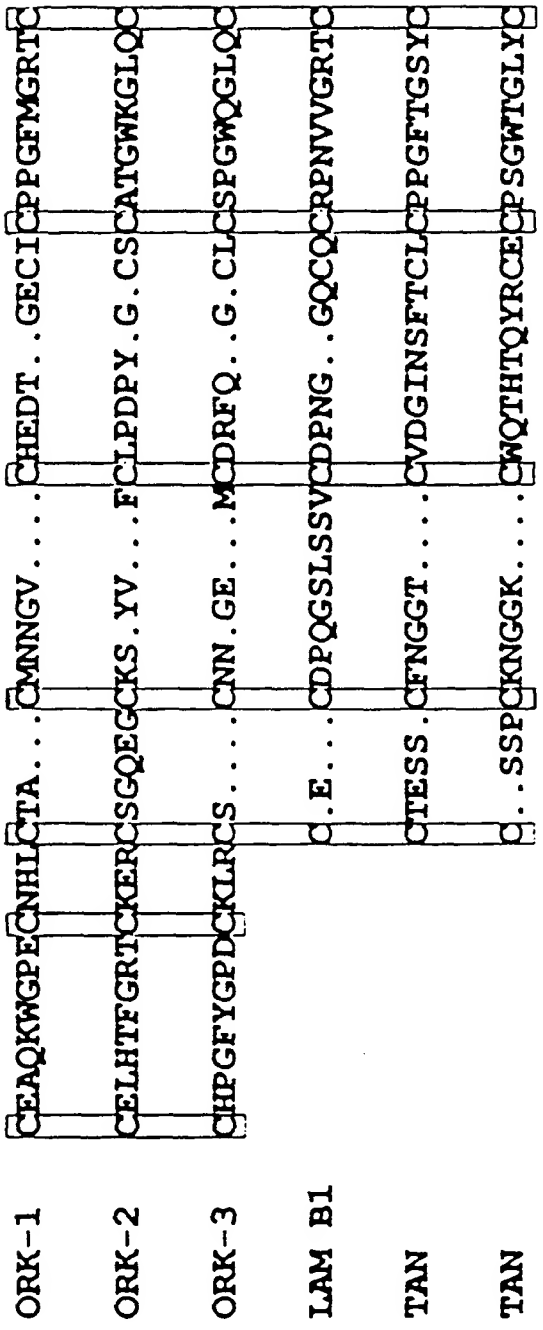


FIGURE 3A

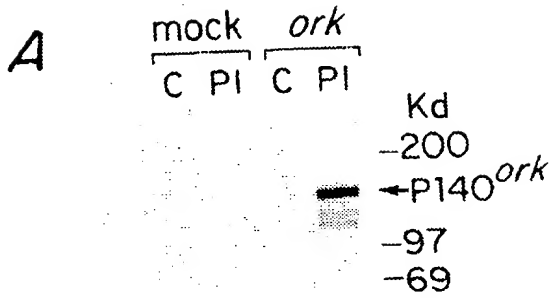


FIGURE 3B

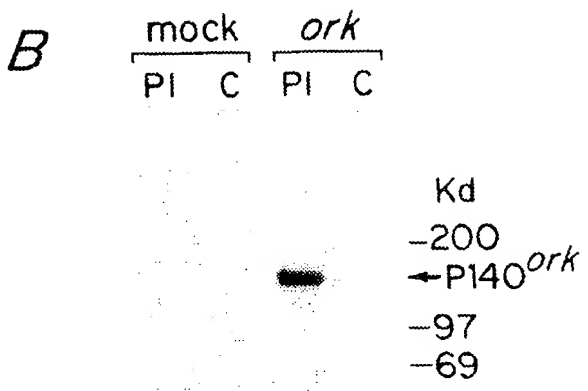


FIGURE 3C

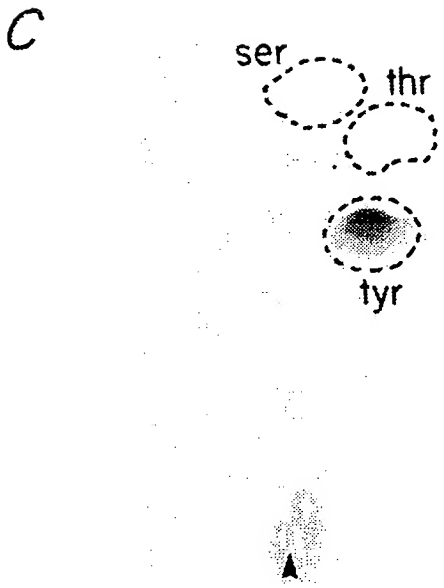


FIGURE 3D

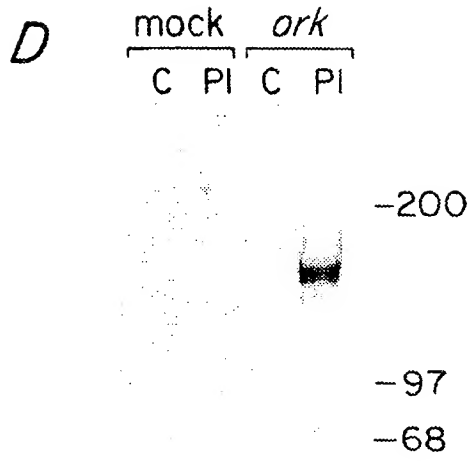


FIGURE 4

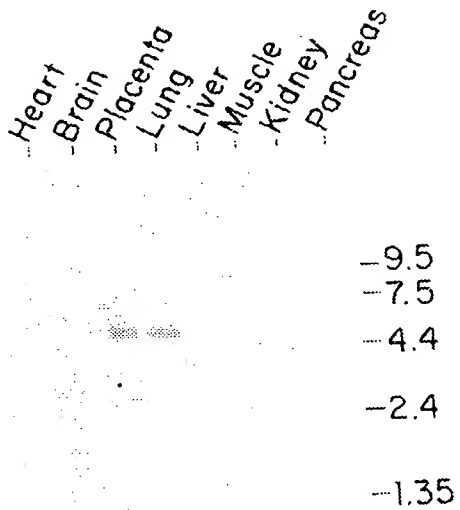


Figure 5a

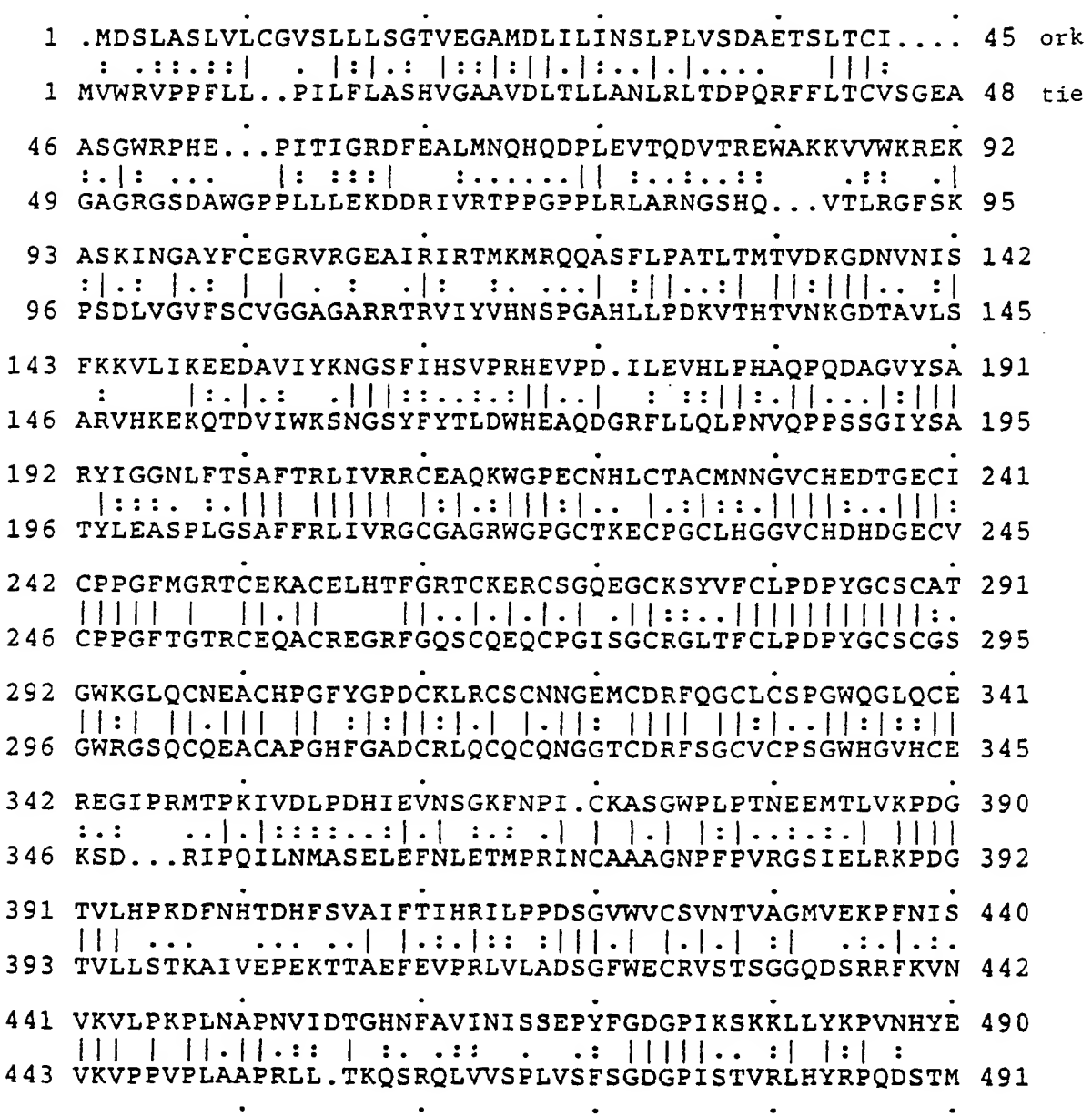


Figure 5b

```
 491 AWQHIQV.TNEIVTLNYLEPRTEYELCVQLVRRGEGGEGHPGPVRRFTTA 539
     .|   | | ..| |||   | |:|:|.: ||| |.|||||| ||.  :||.
 492 DWSTIVVDPSENVTLMNLRPKTGYSVRVQLSRPGEGGEGAWGPPTLMTTD 541

 540 SIG.LPPPRGLNLLPKSQTTLNLTW.QPIFPSS..EDDFYVEVERRSVQK 585
     : : | .|:  .: ..: . |.:.|  |:.|:.  :|:|.: :   . ..
 542 CPEPLLQPWLEGWHVEGTDRLRVSWSLPLVPGPLVGDGFLLRLWDGTRGQ 591

 586 SDQQNIKVPGNLTSVLLNNLHPREQYVVRARVNTKAQGEWSEDLTAWTLS 635
     . .:|:. |.. |. ||..| | .:| : ..:   .  : ... .   |.
 592 ERRENVSSPQARTA.LLTGLTPGTHYQLDVQLYHCTLLGPASPPAHVLLP 640

 636 DILPPQPENIKISNITHSSAVISWTILDGY..SISSITIRYKVQGKNEDQ 683
     .  ||.| ::.  .:..|.  :.|.  ::.  .||...:  .|.| .:|.
 641 PSGPPAPRHLHAQALSDSEIQLTWKHPEALPGPISKYVVEVQVAGGAGDP 690

 684 ...HVDVKIKNATIIQYQLKGLEPETAY......QVDIFAENNIGSSNPA 724
        .|| . ...||| :||::.| |       :: ::: . . .:..
 691 LWIDVDRPEETSTII....RGLNASTRYLFRMRASIQGLGDWSNTVEEST 736

 725 FSHEL...VTLPESQAPADLGGGKMLLIAILGSAGMTCLTVLLAFLIILQ 771
     ::::|   ..:.||.| |: | :. |::|::||.: ||||:| |:|.::
 737 LGNGLQAEGPVQESRA.AEEGLDQQLILAVVGSVSATCLTILAALLTLVC 785

 772 LKRANVQRRMAQAFQNVR.EEPAVQFNSGTLALNRKVKNNPDPTIYPVLD 820
     ::|. ::||.. .:|... ||. :||.||||.|.|:.| .|:|  ||||:
 786 IRRSCLHRRRTFTYQSGSGEETILQFSSGTLTLTRRPKLQPEPLSYPVLE 835

 821 WNDIKFQDVIGEGNFGQVLKARIKKDGLRMDAAIKRMKEYASKDDHRDFA 870
     |:||.|:|:|||||||||::|.|||||:|:||||.:||||.:||||||
 836 WEDITFEDLIGEGNFGQVIRAMIKKDGLKMNAAIKMLKEYASENDHRDFA 885

 871 GELEVLCKLGHHPNIINLLGACEHRGYLYLAIEYAPHGNLLDFLRKSRVL 920
     ||||||||||||||||||||||.:|||||:||||||.||||||||||||
 886 GELEVLCKLGHHPNIINLLGACKNRGYLYIAIEYAPYGNLLDFLRKSRVL 935

 921 ETDPAFAIANSTASTLSSQQLLHFAADVARGMDYLSQKQFIHRDLAARNI 970
     ||||||| .::|||||||.|||:||.|.|.||:|||:||||||||||||:
 936 ETDPAFAREHGTASTLSSRQLLRFASDAANGMQYLSEKQFIHRDLAARNV 985

 971 LVGENYVAKIADFGLSRGQEVYVKKTMGRLPVRWMAIESLNYSVYTTNSD 1020
     |||||...||||||||||:|||||||||||||||||||||||||||.||
 986 LVGENLASKIADFGLSRGEEVYVKKTMGRLPVRWMAIESLNYSVYTTKSD 1035

1021 VWSYGVLLWEIVSLGGTPYCGMTCAELYEKLPQGYRLEKPLNCDDEVYDL 1070
     |||:|||||||||||||||||||||||||||: |:|.| |||||||:|
1036 VWSFGVLLWEIVSLGGTPYCGMTCAELYEKLPQADRMEQPRNCDDEVYEL 1085

1071 MRQCWREKPYERPSFAQILVSLNRMLEERKTYVNTTLYEKFTYAGIDCSA 1120
     ||||||::|||||.|||| : |.||||.||.||| .|:|.|||||||..|
1086 MRQCWRDRPYERPPFAQIALQLGRMLEARKAYVNMSLFENFTYAGIDATA 1135

1121 EEAA* 1125
     |||
1136 EEA.. 1138
```

Figure 6

```
   3 VPWLGACEHRGYLYLAIEYAPHGNLLDFLRKSRVLETDPAFAIANSTASI 52   m
     : :||||||||||||||||||||||||||||||||||||||||||||||||.
 886 INLLGACEHRGYLYLAIEYAPHGNLLDFLRKSRVLETDPAFAIANSTAST 935  h

  53 MSSQQLLHFAADVARGMDYLSQKQFIHRDLAARNILVGENYIAKIADFGL 102  m
     :|||||||||||||||||||||||||||||||||||||:||||||||
 936 LSSQQLLHFAADVARGMDYLSQKQFIHRDLAARNILVGENYVAKIADFGL 985  h

 103 SRGQEVYVKKTMGRLPVRWMAIESLNYSVYTTNSDVWSYGVLLWEIVSLG 152  m
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 986 SRGQEVYVKKTMGRLPVRWMAIESLNYSVYTTNSDVWSYGVLLWEIVSLG 1035 h

 153 GTPYCGMTCAELYEKLPQGYRLEKPLNCDDEVYDLMRQCWREKPYERPSF 202  m
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1036 GTPYCGMTCAELYEKLPQGYRLEKPLNCDDEVYDLMRQCWREKPYERPSF 1085 h

 203 AQILVSLNRMLEERKTYVNTTLYEKFTYAGIDCSAEEAA 241  m
     ||||||||||||||||||||||||||||||||||||||||
1086 AQILVSLNRMLEERKTYVNTTLYEKFTYAGIDCSAEEAA 1124 h
```

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/06093

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C07H 21/00; C07K 15/00; C12N 9/12, 15/15, 15/54, 15/63; C12P 21/02
US CL :435/69.1, 194, 320.1; 530/387; 536/23.2, 24.5
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 69.2, 194, 320.1; 530/350, 351, 387, 395, 396, 399; 536/23.2, 24.5; 935/9, 13, 14, 34, 36

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	Oncogene, Volume 8, issued 1993, S.F. Ziegler et al., "Molecular Cloning and characterization of a novel receptor protein tyrosine kinase from human placenta", pages 663-670, entire document.	1-20
X	US, A, 4,543,439, (Frackelton et al.) 24 September 1985, entire document.	19
X,P Y	Oncogene, Volume 7, issued 1992, D. L. Dumont et al., "tek, a novel tyrosine kinase gene located on mouse chromosome 4, is expressed in endothelial cells and their presumptive precursors", pages 1471-1480, especially figure 1.	13,17-18 1-12, 14-16, 19-20
Y	Cell, Volume 63, Issued 05 October 1990, J.G. Flanagan et al., "The kit ligand: a cell surface molecule altered in steel mutant fibroblasts", pages 185-194, entire document.	2, 6, 10, 15

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be part of particular relevance
"E" earlier document published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion f the international search
17 September 1993

Date of mailing of the international search report
SEP 1993

Name and mailing address of th ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. NOT APPLICABLE

Authorized fficer
ROBERT A. WAX
Telephon No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	Proc. Nat. Acad. Sci, U.S.A., Volume 87, issued November 1990, J. Partanen et al., " Putative tyrosine kinases expressed in K-652 human leukemia cells", pages 8913-8917, entire document.	1, 5, 13-14 2, 6, 9-10, 15
X Y	Molecular and Cellular Biology, Volume 12, Number 4, issued April 1992, J. Partanen et al., "A novel endothelial cell surface receptor tyrosine kinase with extracellular epidermal growth factor homology domains, pages 1698-1707, entire document.	1, 5, 9, 13-14 2, 6, 10, 15
Y	Molecular and Cellular Biology, Volume 11, Number 10, issued October 1991, J. P. O'Bryan et al., "axl, a transforming gene isolated from primary human myeloid leukemia cells, encodes a novel receptor tyrosine kinase", pages 5016-5031, entire document.	9-12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/06093

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, Dialog ~~files 155, 5, 434~~ (Medline, Biosis, Scisearch), PIR, EMBL, Genbank
search terms: tyrosine kinase??, receptor??, genetics, DNA, orphan, ork, tie, tek, express?, solub?

INTERNATIONAL SEARCH REPORT

Inte...tional application No.
PCT/US93/06093

Box I Observations where certain claims were f und unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because ~~they relate to~~ subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
(Telephone Practice)

I. Claims 1-18, drawn to recombinant DNAs encoding ork, and their expression products, classified in 435/69.1.
II. Claim 19, drawn to an antibody immunoreactive with ork, classified in 530/387.1.
III. Claim 20, drawn to inhibitory oligonucleotides, classified in 536.24.5.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark n Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.